Doctoral Thesis

Realistic Microscopy Image Translation using Multi-Task Learning and Structure-Aware Constraints for Label-Free High-Content Screening

GyuHyun Lee

Department of Computer Science and Engineering

Ulsan National Institute of Science and Technology

2021

Realistic Microscopy Image Translation using Multi-Task Learning and Structure-Aware Constraints for Label-Free High-Content Screening

GyuHyun Lee

Department of Computer Science and Engineering

Ulsan National Institute of Science and Technology

Realistic Microscopy Image Translation using Multi-Task Learning and Structure-Aware Constraints for Label-Free High-Content Screening

A thesis/dissertation submitted to
Ulsan National Institute of Science and Technology
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Gyuhyun Lee

12/09/2020 of submission

Approved by

Advisor

Se Young Chun

Realistic Microscopy Image Translation using Multi-Task Learning and Structure-Aware Constraints for Label-Free High-Content Screening

GyuHyun Lee

This certifies that the thesis/dissertation of GyuHyun Lee is approved.

12/09/2020 of submission

Signature

Advisor: Se Young Chun

Signature

Jae-Young Sim: Thesis Committee Member #1

Signature

Kwang In Kim: Thesis Committee Member #2

Signature

Sungahn Ko: Thesis Committee Member #3

Signature

Won-Ki Jeong: Thesis Committee Member #4

Abstract

Image processing is an important and inevitable pipeline for a wide range of industies such as autonomous car, manufacturing, search engine, and healthcare, to solve existing problems. Recently, the image processing in health care has focused on developing more precise prediction methods due to advancements in computing speeds and deep learning methods. Even though all of these challenges are non-trivial in the biomedical field, the most important issue among the existing problems is to determine proper targeted cancer therapy for individual patient in order to achieve precision medicine. Specially, progressing to the rapid acquisition times necessary to generate plenty of microscopy images for biomedical samples, which cannot be observed with the naked eye. Based on these microscopy images, various drug responses to patient-derived cell cultures can be analyzed by stained individual cells with various biomarkers to gain a more detailed understanding through high-content screening (HCS). In this dissertation research, several novel image translation contributions for achieving software-based HCS for precision medicine.

First, a novel image translation method, *DeepHCS*, for transforming bright-field microscopy images into synthetic fluorescence images of cell nuclei biomarkers is introduced. The main motivation of the proposed work is to automatically generate virtual biomarker images from conventional bright-field images, which can greatly reduce time-consuming and laborious tissue preparation efforts and improve the throughput of the screening process. DeepHCS uses bright-field images and their corresponding cell nuclei staining (DAPI) fluorescence images as a set of image pairs to train a series of end-to-end deep convolutional neural networks.

Second, a novel microscopy image translation method is proposed, $DeepHCS^{++}$, for transforming a bright-field microscopy image into three different fluorescence images to observe apoptosis, nuclei, and cytoplasm of cells, which visualize dead cells, nuclei of cells, and cytoplasm of cells, respectively. Thus, the main contribution of the proposed work is the automatic generation of three fluorescence images from a conventional bright-field image using multi-task learning with adversarial losses; this can greatly reduce the time-consuming and laborious tissue preparation process as well as improve throughput of the screening process. $DeepHCS^{++}$ uses multi-task learning with adversarial losses to generate more accurate and realistic microscopy images.

Third, an image translation method with structure-aware features is proposed for the acquisition of more realistic fluorescence microscopy images. This method integrates multi-task learning and cyclic consistency. In order to attain such realistic microscopy images, this proposed method employs an autoencoder that generates cell profile feature maps, in which include satisfactory cell textures and revise feature maps from the translation network by cooperating with the mixture network over these two different feature modalities.

Contents

List of Figures

List of Tables

List of Algorithms

I Introduction

The importance of developing drug discovery is even recognized by the general public, who did not study medicine professionally. The most important purpose of new drug development is to improve human health and treat new diseases. Recently, the social environment, such as aging, has changed, and rare diseases that were previously undiagnosed are now continually being discovered due to the advancements in diagnostic technology. The importance of developing new drugs has been increasing day by day, especially at a time when infectious diseases such as COVID-19 are ravaging the world. The development of drug discovery, such as proper vaccines and treatments, is actively underway and most people agrees with its importance.

Depending on the research institution, to rapidly develop new drugs, the acceleration of high-speed multi-scanning systems analyzing tens of thousands to millions of candidate substances (chemicals) is key. As a result, the screening process is costly and time-consuming and requires an environmental configuration that is efficient and capable of screening large numbers of candidate chemicals in a short period of time. Many researchers leverage automated high-throughput screening (HTS) systems to quickly and efficiently handle this set of processes. HTS is a method for rapidly testing thousands of compounds in parallel for their activity in one or more biological assays. As such, HTS has been a key process in drug discovery allowing the automated testing of chemical activity in a specific biological target for many years [3, 4]. In HTS, a large number of compounds being considered for specific biological targets are injected to analyze and confirm their reactive responsiveness by quantifying the activation of certain proteins via the activation of fluorescence. Due to the HTS system, screening of more than 10,000 compounds is enabled within a day. A series of experimental processes measuring the activities of various proteins associated with this particular biological target can be applied to robotic systems to save enormous researcher and screening times. Therefore, HTS is not just a high-speed screening technology, but is considered a key factor and the most fundamental technology for successful screening during the development of new drugs.

Although many pharmaceutical companies and research institutes have invested heavily in existing HTS technologies, the primary screening process has been largely considered to reduce the huge economic costs incurred during the secondary screening process, which is an optimization process for candidate chemicals discovered in the initial screening stage. For this issue, high-content screening (HCS) [5] was introduced and is considered a complex and functional screening technique based on highly-sensitive fluorescence images with varying temporal or spatial resolution of the internal compartment of cells. HCS is able to obtain the degree of fluorescence of effective materials on biological targets by microscopy is especially advanced in optical technology, and can simultaneously detect the fluorescence of multiple wavelengths to image proteins in cells. This enables a more in-depth analysis of cell mechanisms through the observation of morphological features of individual cells as well as reactive analysis.

In my dissection, I aim to further enhance the automation and acceleration of screening

techniques in the image data processing of HCS by proposing image processing methods with state-of-the-art deep learning techniques, thereby increasing the efficiency of screening techniques and significantly improving the stability of new drug development.

1.1 Motivation

HCS technologies are evolving day by day, thanks to advanced microscopic technologies and high-speed computing technologies as described previous section. HCS is characterized by the ability to image the fluorescence reactions of cells through a microscope. Among the different types of microscopes, the fluorescent microscope was used to photograph custom fluorescent microscopy images for analysis and conduct individual cell analysis in my dissertation. The process of creating these fluorescence images is not simple. Cell culture is carried out to directly detect biological tissues related to disease in real patients and be fully utilized in experiments. Several biomarkers are then dyed in cells and stuffed with cells in order to capture fluorescence images for analysis purpose. Fluorescence images of cells were then photographed using a fluorescent microscope. However, all cells which were taken during the process of dyeing and stuffing ultimately die. Fluorescence images represent the cell's properties by dyeing its nucleus or cell state. In contrast, bright-field images can take a picture of the actual cell itself to identify the entire cell form; however, the quality of the image varies greatly depending on the environment of the photograph. Fluorescence images that can be acquired through dyeing and stuffed in cells and those that can be acquired in simple photography will have completely different characteristics and be used for different purposes.

However, bright-field images have the ability to acquire a photographic record of cells without any extra preparation, while fluorescence images require time-consuming cell fixation and staining procedures. Another advantage of the bright-field image is its ability to capture the dynamics of cells because cell fixation and cell staining are not required. However, fluorescence imaging can capture only a snapshot of the cells at any given point in time because cells die during fixation and staining. Due to the advantages of these bright-field images, many studies have conducted much of research on cell analysis using bright-field images. However, in real-world industrial and research activities, analysis is carried out with consideration of fluorescent microscopes for drug response and individual cell analysis.

The primary motivation of this dissertation research comes from the following observations. The simplest image obtained by photographing cells using an electron microscope is a bright-field microscope image. Bright-field microscopy images allow the cell's morphology to be observed by imaging the cells in the well as they do not require any pre-processing. Cells are living organisms and, thus, do not remain stationary but continue to move as long as they are alive. Therefore, by observing the captured bright-field image, it is easy to see that most of the cells are overlapped or attached (Fig. 1). In the case of a fluorescence microscope, which captures photographs with light from below, cells are frequently imaged faintly depending on the photographing setting. Nevertheless, many studies [6–8] have been actively conducted to analyze individual cell images

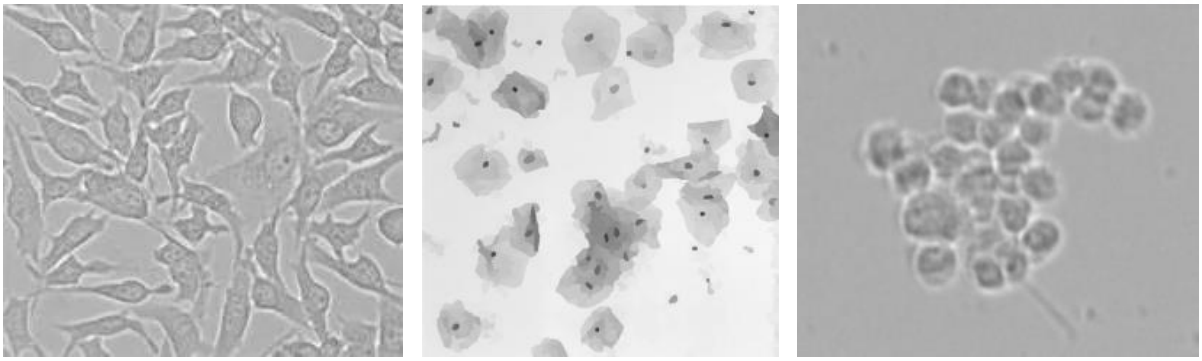


Figure 1: Examples of touched, overlapped, and blurred cells in bright-field images.

from easily obtainable bright-field images. Although much research has been conducted and remarkable results have been achieved, it can all be easily acquired. This is due to the quality of bright-field images that change with each shot. Therefore, in the industry of precision medicine, cells are stained using fluorescent biomarkers, which are then imaged and analyzed.

1.2 Problem Statement

As mentioned before, HCS is also used significantly in drug discovery and anticancer treatment. In terms of new drug development and the COVID-19 pandemic, many pharmaceutical companies and medical researchers have developed fast-track drug discovery, and their efforts can be summarized into three main strategies: the repurposing of existing drugs, development of new therapeutics, and development of vaccines. In order to accelerate this process, they can leverage existing technologies, including high-content screening (HCS), which is already well-established as playing a significant role in viral disease research and drug discovery.

From a different perspective, i.e., the point of view of chemotherapy, the best treatment options for many cancers are currently surgery and chemotherapy approaches and it is the best treatment option now. For example, glioblastoma (GBM), which is a brain tumor commonly found in the cerebral hemisphere of the brain but can be widely found in the other central nervous regions such as spinal cord, accounts from 12% to 15% of all brain tumors. GBMs are considered fatal brain tumor because it could not be entirely treated, and the limited lifespan not increasing despite medical techniques having progressed immensely in a few decades [9]. When patients are diagnosed with GBM, in most cases, the best treatment option as of today is surgery to eliminate as many tumor cells as possible because there is no accepted standard treatment yet for individual patients. After the surgery, biologists analyze patient-derived GBM tumor cells to help clinicians prescribe the proper chemotherapy for patients. In this case, FDA-approved drugs or targeted drugs are screened to determine which drug is the most effective for the specific patient. However, the surgery can highly risky and lead to complications [10]. Based on this issue, in addition to surgical treatments, patient-specific chemotherapy by analyzing patient-driven GBM tumor cells to find the most effective drug for the target patient, called *precision*

medicine [11, 12], has become popular.

Based on these two perspectives, the high-speed of HCS is becoming an even more critical factor. However, it was determined that in the process of dying and stuffing cells, an essential step in obtaining fluorescence images, were still bother some stages in high-speed HCS. Therefore, my dissertation assumes that high-speed screening could be achieved with fluorescence images converted from bright-field images to purpose-built ones.

The goal of this dissertation research is to develop an intelligent image translation technology capable of acquiring fluorescence microscope images necessary to meet the needs in a real industrial field from a bright-field microscope image. It includes the study on developing fluorescence microscopy image translation technology specialized for bright-field microscopy images and verifying whether the image conversion technology can sufficiently replace the existing pipeline. In addition, it shows that it can be applied to various cells by verifying various types of cells.

1.3 Research Statement

If various modalities of fluorescence microscopy images suitable for their purpose can be generated effectively and quickly using image processing methods with deep learning instead of physical steps, it is expected that significant savings to society can be expected.

The work proposed in this dissertation dramatically reduce the essential physical steps in conventional HCS workflow that exist between bright-field and fluorescence images by leveraging image processing methods with deep learning to achieve software-based HCS workflows. I introduce image translation methods using state-of-the-art deep learning techniques for converting to multiple fluorescent images with a single bright-field microscope image acquired by autofocusing. First, single-task learning is conducted through deep learning to generate DAPI fluorescence images essential for drug responsiveness. The generated DAPI fluorescence images are analyzed and correlated with the generated fluorescence and real fluorescence images, such as cell viability and IC50, to verify that they can replace the actual DAPI fluorescence images. In-depth analysis of individual cells is carried out through various fluorescence images as well as DAPI fluorescence images. To this end, it is possible to produce several fluorescent images simultaneously with a single microscope image. To this end, multi-task learning will be adopted, and adversarial learning will be used together to obtain more realistic images than before. Cell viability, IC50 analysis is conducted to verify of the generated fluorescence images and morphological analysis of individual cells is also performed. Additionally, to show applicability to multiple cells, I conducted image translations of different cancer cells and analyzed the same drug responsiveness and individual cell analysis to determine the correlation. Furthermore, I present a method to effectively reduce bloated networks and maintain the quality of transformed images to generate multiple fluorescence images.

1.4 Contributions

This dissertation research aimed to develop a novel image translation method using deep learning for software-based high-content screening workflow. The main contributions of this dissertation research are below.

- **Bright-Field to DAPI Fluorescence Microscopy Image Conversion for Drug Response in Label-Free High-Content Screening**
  I propose a novel image translation method from a bright-field image to DAPI fluorescence image that includes core information on cell nuclei and is essential in conducting drug response tests on cells. In order to develop this method, the shape of the U-net network architecture was defined, and a residual block was applied to define deeper neural networks. Two defined encoder-decoder structures were connected in succession. The front network was used for the purpose of image conversion, and the latter network was used for the purpose of modifying the translated image again. At this time, the network with the purpose of modification takes a clue in making corrections by providing both the bright-field image and the translated image as inputs. In addition to comparing the translated image with the real fluorescence image through the proposed method, the similarity is verified by comparing the drug response analysis in the real fluorescence image with the analysis result in ours.

- **Bright-Field to Multiple Fluorescence Microscopy Image Conversion for Individual Cell Analysis in Label-Free High-Content Screening**
  I propose a novel end-to-end image translation method from a bright-field image to multiple fluorescence images for drug response as well as morphological analysis of individual cells. I employed multi-task learning to concurrently generate multiple biomarkers and adversarial losses [13], which allowed for the generation of more realistic fluorescence microscopy images. I also tested the proposed method on three time-lapse data (each consisted of 65 time-lapse bright-field images). The proposed method effectively avoided the time-consuming and laborious tissue preparation process for generating biomarkers while providing accurate image analysis results using the conventional image-based HCS workflow.

- **Structure-Aware Microscopy Image Translation using Autoencoder for Cell Profile Feature Maps**
  I propose an efficient and lightweight image translation method to reduce the number of trainable parameters and generate more realistic fluorescence images than previous works. For this proposed method, the autoencoder is employed to generate cell profile feature maps used as reference information, such as cell shape. These cell profile feature maps are compared with feature maps from the translation network via a cyclic consistency. I evaluated the accuracy of the proposed method using widely used image quality metrics, PSNR

and SSIM, and validated qualitatively between our results and real cytoplasm fluorescence images using error map visualization and R^2 correlation.

1.5 Document Organization

In Section II, I explain the background knowledge on the high-content screening and multiple modalities fluorescence images generated by the high-content screening. In Section II, I introduce the high-content screening and previous work on conventional and deep learning method for image processing for medical images. In Section IV, I explain my research works to solve by replacing the conventional high-content screening with software-based high-content screening using the deep learning method. In Section V, novel multi-task learning to generate multiple fluorescence images from a bright-field image is discussed. In Section VI, an efficient fluorescence image translation using the autoencoder and cyclic consistency not used in previous methods are introduced to reduce the number of trainable parameters. Finally, Section VII wraps up the my dissertation and suggests possible future research directions.

II Background and Related Work

2.1 Overview of High-Content Screening

HCS has been used as an auspicous and promising analysis method for new drug development as an automated, high-speed screening method since its release more than 20 years ago [5, 14–16]. It is widely used not only in pharmaceutical companies but also in universities and research institutes, and is used for the development of new drugs and in other biological and medical fields.

HCS is a methodology for systematically discovery phenotypes from image data in cellular assays [17]. In contrast to traditional HTS, which has a single readout of activity, HCS allows researchers and scientists to measure many features of individual cells at once. HTS is an experimental setup for systematically testing many experimental conditions. This is typically achieved using a very simple readout of, for example, cell viability or a univariate measure of cytotoxicity. HCS performs these kinds of screens utilizing a more complex and informative readout by leveraging multiple fluorescence microscopy channels. HCS thus increases the content of the readout to a missive number of features while maintaining the throughput. HCS has been instrumental in decoding the molecular basis several diverse biological processes, such as cell division [18], protein secretion [19]. HCS has also been used to systematically screen for the localization of biomolecules inside cells [20]. HCS also plays a role in the early stages of the drug discovery process [21, 22] in which a cell line population is exposed to small molecules of drug compounds.

In the HCS process, cell fluorescence images can generally be acquired through several steps, as shown in Fig. 2. In Fig. 2, the first step is to obtain a sample by taking cancer tissue from the actual patient's cancer area. This sample is cultured for sufficient experimentation by proliferating the cells. The proliferated cells are then distributed in a plate consisting of as few as six to as many as 386 wells, and various drugs are injected at different doses. After an certain amount of time, a specific biomarker is injected into each well to stain the nucleus or cytoplasm of the cell, and the cells are stuffed using bovine serum for imaging. After these processes, all preparations for imaging are completed. When cells are ready to be photographed, various types of electron microscopes can be used. The microscope used in my research is specifically used as a transmitted-light microscope (the fluorescence microscope, Fig. 3). Each well is imaged at a specific magnification through a transmitted-light microscope to acquire bright-field microscopy and corresponding fluorescence microscopy images. The acquired fluorescence images enables in-depth analysis of the morphological changes within cells according to the fluorescence response of various compounds (IC50 and AUC). HCS workflow allows the acquisition and storage of various fluorescence images reflecting different characteristics of cells, which has the great advantage of continuing to analyze these stored images according to their purpose. However, the disadvantage of this comes from the storage needed for the amounts of image data acquired through large-scale

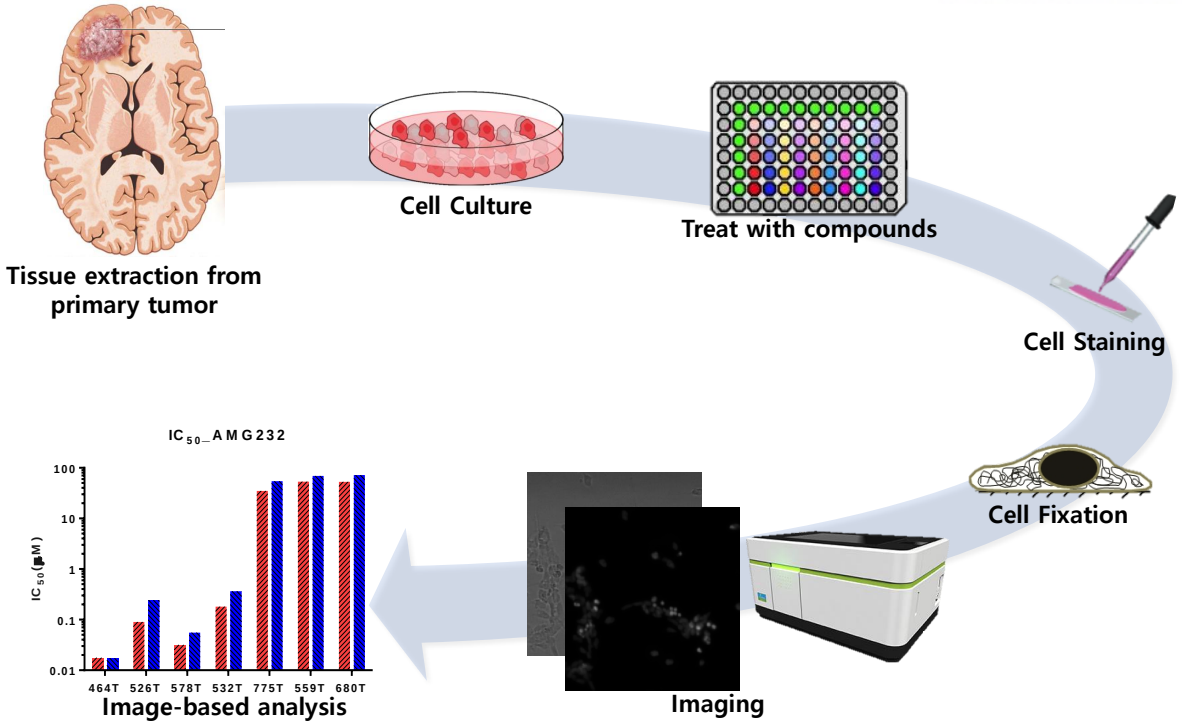


Figure 2: The overview of the HCS pipeline from the tissue extraction to the image analysis for the precision medicine.

screening, which can incur considerable costs [23].

Although HCS is in the spotlight as a high-speed screening method compared to past methods; many steps are taken from the acquisition of fluorescence images to the start of image-based analysis, as shown in Fig. 2. The actual imaging process is still relatively time-consuming. It takes several hours per 386-plate to capture bright-field images and fluorescence images of cytoplasm and nuclei even when the commonly used 386-plate is specified at 10× magnification. In order to deal with this issue, the field of virtual screening has been proposed [24], and recently, in-silico label methods [25, 26] have been proposed to acquire fluorescence images directly from bright-field images for different purposes.

2.2 Different Modalities of Microscopy Fluorescence Image

In HCS, imaging data derives from one or another of the many types of microscopy. Fundamental to high content screening is fluorescence microscopy, which uses a laser to excite fluorescent molecules in organic matter. These molecules are known as fluorophores and emit light at a unique wavelength (color) upon excitation by a laser. As such, localized fluorophores can be utilized to highlight key cellular regions. The predominant technique used in HCS is immunofluorescence, which relies on fluorescently-labeled antibodies. In most screening applications, the nuclei are stained with one of these techniques, allowing for the subsequent identification of individual cells. The other fluorescent markers are selected in accordance with the research needs,

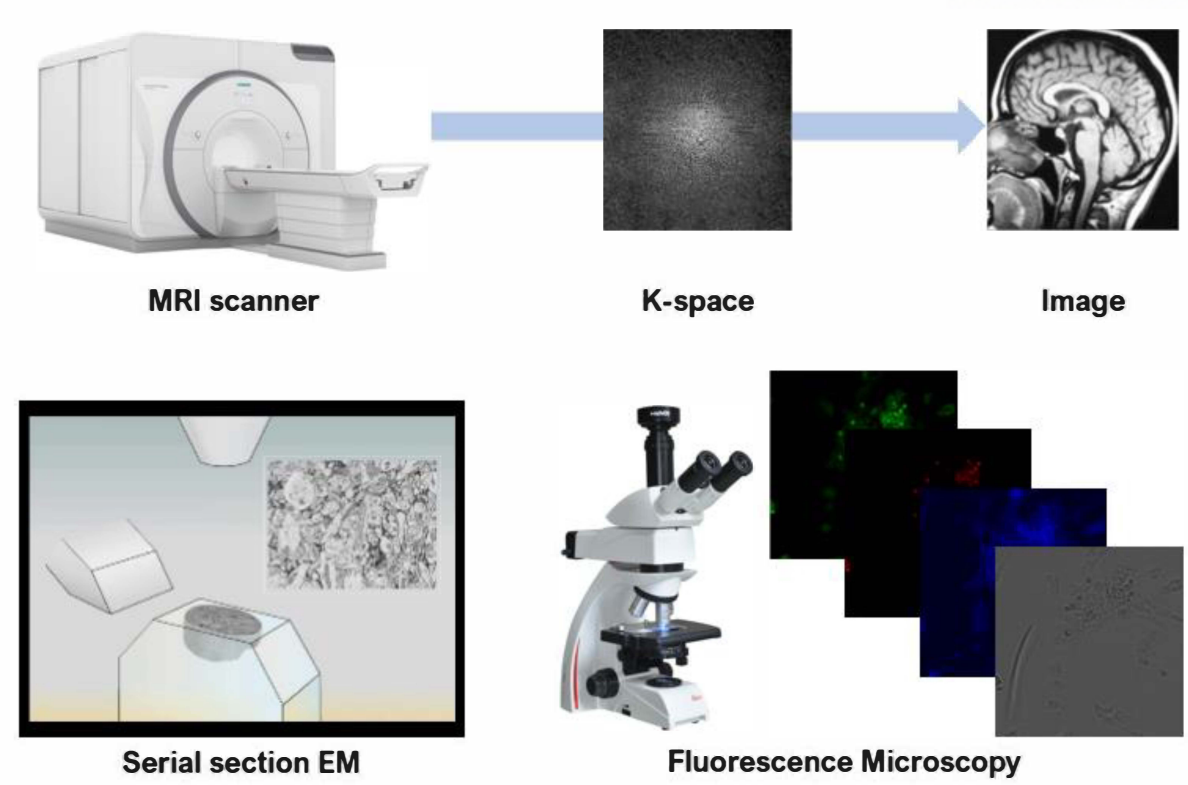


Figure 3: Examples of various acquisition method to generate different modalities bio-medical images.

for example, apoptosis, or cytoplasm. Thus, the fluorescence markers of the screen, responding to distinct wavelengths of light, yield a set of multiplexed images painting a composite picture of the cell in its critical sub-cellular structures. As mentioned before, all image data used in all experiments of my research is imaged by the fluorescence microscope (Fig. 3).

Bright-field microscopy (Fig. 4a) passes visible light through a sample, producing a picture in which light is attenuated according to the varying densities of the imaged specimen.

Apoptosis fluorescence images (Fig. 4b) represent dead cells stained with Alexa 488. The used biomarker only reacts to the dead cell nucleus, and if we observe the apoptosis fluorescence images, the dead cell nucleus has a high pixel value; otherwise, it has a low pixel value, so, we can observe evidently dead cells.

DAPI fluorescence images (Fig. 4c) represents cell nuclei stained with Hoechst 33342. The used biomarker stains all of the cell nuclei, and the dead cell nucleus has a high pixel value while the live cell nucleus has a low pixel value. Due to thjis distinct feature, the cell viability [27] is analyzed in DAPI fluorescence images for drug response or drug discovery.

Cytoplasm fluorescence images (Fig. 4d) represents the cell cytoplasm stained with Alexa 594. The used biomarker responds to the entire cell, which is evaluated for morphological features such as the length, width, roundness, etc.

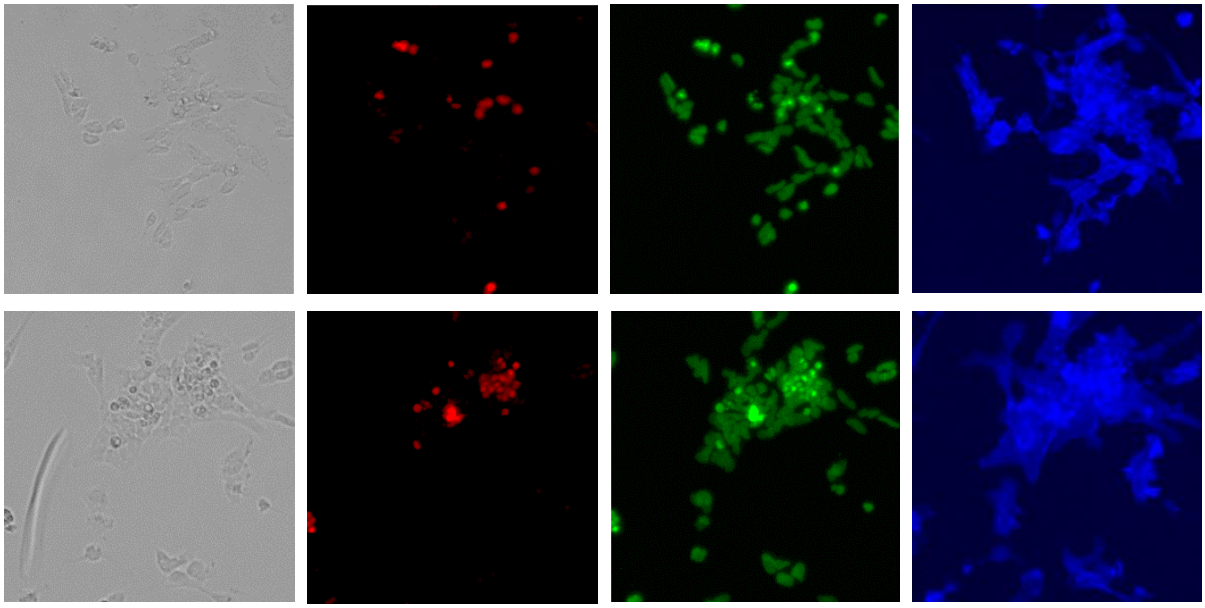


Figure 4: Examples of bright-field and the corresponding fluorescence images. Each row represents a different region of interest: (a) bright-field images captured using microscopy without any biomarker; (b) apoptosis fluorescence images stained with Alexa 488 for detecting dead cells; (c) DAPI fluorescence images stained with Hoechst 33342 for cell nuclei; and (d) cytoplasm fluorescence images stained with Alexa 594 for cell morphology.

III Related Work

3.1 Bright-Field Microscopy Image Processing

Most previous notable studies focused on leveraging multi-focal bright-field images in the analysis pipeline to detect cell structures with minimal to no support from fluorescence images. Selinummi *et al.* [28] used multi-focal bright-field images to extract the shape of cells without whole cell fluorescence images. They calculated the intensity variation along the z-stack of multi-focal bright-field images to robustly detect cell boundaries. Ali *et al.* [29] proposed the detection and segmentation of adherent HT1080 and HeLa cells using bright-field images. This method extracted the local phase and orientation from multi-focal bright-field images using a monogenic single framework to guide the evolution of the active contour. Tikkanen *et al.* [30] employed a machine learning approach using a histogram of oriented gradient (HOG) feature [31] for detecting cells in 25 focal bright-field images. The extracted features and their neighboring intensity histograms were combined for classification using a support vector machine [32]. Limatainen *et al.* [33] employed a logistic regression with a ℓ_1 penalty to classify the location of cells and non-cells using the intensity values from 25 focal images as features.

3.2 Image-to-Image Translation

Since the pioneering work of [13], using generative adversarial networks to generate realistic images has becomes popular. These methods include texture synthesis [34], image translation [1, 35–37], etc. DiscoGAN [36] used an auto-encoder [38] to automatically determine the relationship between two visual domains without any explicit pair labels. Unlike a standard GAN [13], the generator takes the input image from one domain and produces a fake image; whereas, the discriminator takes a fake image from the generator and real image from another domain to distinguish real and fake. CycleGAN [35] learned to translate from a source domain to a target domain without a paired training set. This method introduced cycle-consistency to map one domain to another domain; their model is similar to DiscoGAN [36]. UNIT [37] proposed mapping the shared-latent space by utilizing encoding functions from two different domains. This method used a variational auto-encoder [39]; the core features of two different domains are in shared-latent space. Two generation functions can then, based on the shared-latent space, generate the image for their purpose. Pix2Pix [1] studied conditional adversarial loss. Their process involved concatenating both an input image and a translated image and then feeding its concatenated images into a discriminator. Similar to the above work, image translation is also applied to the microscopic image processing phase. Han *et al.* [40] proposed a translation from the differential interference contrast (DIC) image to phase contrast image, and vice versa. In this method, a cell mask [41] was additionally used to restore the invisible region and give the generator more textural information because DIC images produced transparent cell structure results. This method is an application of image translation microscopy using adversarial loss. Even though this can translate different microscopy images, the performance of this approach is not superior to that of state-of-the-art techniques. Fu *et al.* [42] generated synthetic binary volumes from subvolumes of the original volume and translated synthetic 3D volumes into synthetic binary volumes by leveraging spatially-constrained CycleGAN with 3D convolutional layers. For spatial constraint, an additional generator was employed to eliminate spatial shifting at the location of the nuclei between the synthetic 3D volume and binary volume. These pairs of synthetic 3D volumes and corresponding binary volumes were utilized as the next step for 3D segmentation via 3D U-Net architecture.

3.3 Multi-Task Learning

Multi-task learning has achieved good performance across many research fields of machine learning, from natural language processing [43, 44] and speech recognition [45] to computer vision [46–48]. R. Caruana *et al.* [49] explained that multi-task learning is an approach to domain transfer between multiple tasks that improve generalization using the domain information in the training signals of related tasks. It involves using a domain transfer to leverage additional information sources to improve learning performance on the current task. The reasons multi-task learning works well are explained by Ruder *et al.* [50]: implicit data augmentation,

attention focusing, eavesdropping, representation bias, and regularization. Due to these advantages, multi-task learning can achieve better performance than single-task learning when tasks are related and share common knowledge.

3.4 In-Silico Labeling

More recently, several methods have been proposed for predicting fluorescence images from transmitted-light images such as bright-field and phase images [51]. Transmitted-light images were used to predict fluorescence images [52]. They used transmitted-light z-stacks consisting of out- and in-focus images for input data and generated several fluorescence images using the deep learning method. They generated several modalities of fluorescence images; however, the acquisition of transmitted-light images might be time-consuming as this method required at least 13 stacks to be processed at the same time as the input data. Another method [26] employed the U-net architecture [53] to predict corresponding fluorescence images from 3D transmitted-light images and 2D electron micrograph images. A 3D convolutional layer was applied to the former images, whereas a 2D convolutional layer was used for the latter images. This method received 32 stacks as input data and generated 13 different corresponding fluorescence images from 3D transmitted light images. However, certain predicted fluorescence images could not be used for further analysis because the comparison of ground-truth and predicted images from several subcellular structures in all predictions was not relevant with respect to pixel intensities in each image. Li *et al.* [54] proposed unsupervised content-preserving image transformation such as fluorescence restoration, whole-slide histological coloration, and fluorescent labeling. In fluorescent labeling, they translated 13-stack bright-field images to three multi-channel fluorescence images for axons, dendrites, and nuclei only using CycleGAN [35] with a saliency, constraint to separate the background and foreground. Fluorescence images generated using their proposed method do not outperform the conventional supervised scheme, even though the unsupervised approach is attractive. Even though the above methods showed the possibility of converting biomarker images using a computational method, no existing methods were specifically targeting and validated their applicability to high-content screening applications as in our work.

IV Bright-Field to DAPI Fluorescence Microscopy Image Conversion for Drug Response in Label-Free High-Content Screening

4.1 Introduction

GBM is a brain tumor that is commonly found in the cerebral hemisphere of the brain. GBM is considered an obstinate brain tumor because even after medical advances in the past few decades, no effective treatment has been discovered that greatly improves life expectancy in patients. When patients are diagnosed with a GBM, in most cases, the best treatment option is surgery to eliminate as many tumor cells as possible. In addition to surgical treatments, patient-specific chemotherapy by analyzing patient-driven GBM tumor cells to find the most effective drug for the target patient, called *precision medicine*, has become popular. High-throughput screening (HTS) and high-content screening (HCS) have demonstrated their effectiveness in precision medicine in recent studies [11, 55]. Both approaches for precision medicine involve readouts of various drug responses to patient-derived cell cultures. Among them, HCS uses high-throughput imaging and automatic image analysis to evaluate changes in the phenotype of the whole cells, such as counting the number of living cells versus dead cells, measuring the size of the cells, comparing the shape of the cells, etc. In HCS, multiple imaging modalities are commonly used together to image various aspects of the cell phenotypes (Fig. 5). Such imaging modalities include bright-field and fluorescence microscopy, in which the former can capture the overall morphology of the cells, while the latter can image various fluorescent biomarkers. One advantage of using bright-field images in HCS is its ability to acquire a photographic record of cells without any extra preparation while fluorescence images require time-consuming cell fixation and staining procedures. Another advantage of the bright-field image method is its ability to capture the dynamics of cells because cell fixation and cell staining are not required (Fig. 5 lower row). However, fluorescence imaging can capture only a snapshot of the cells at any given point in time because cells die during fixation and staining (Fig. 5 upper row).

Unlike most existing methods that directly analyze bright-field images, DeepHCS converts bright-field images to fluorescence images as accurately as possible using end-to-end convolutional neural networks. By doing this, DeepHCS effectively avoids the time-consuming and laborious cell preparation process for generating biomarkers while providing accurate image analysis results by using the well-established conventional HCS workflow (Fig. 5 bottom row). We evaluate the accuracy of DeepHCS using widely used image quality metrics (e.g., PSNR and SSIM). In addition, we compare cell viability [27], the area under curve (AUC) and the IC50 of the results and real DAPI images to demonstrate that DeepHCS can replace the tissue preparation and fluorescence imaging process in the conventional HCS workflow with the software-based image conversion process.

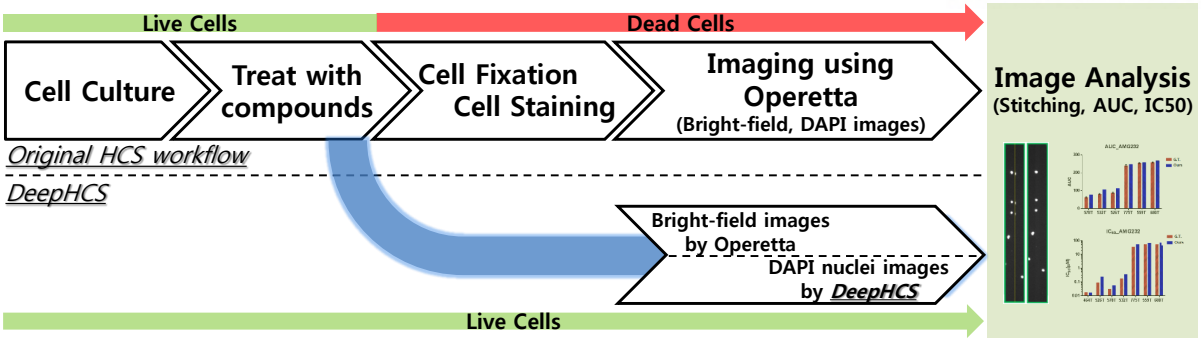


Figure 5: DeepHCS eliminates the cell fixation and staining progress in the original HCS workflow and generates corresponding fluorescence image based bright-field image by Operetta. DeepHCS can keep the cells alive during the entire progress.

4.2 Data

We acquired the image data from patients who had been diagnosed with a GBM brain tumor. The GBM tumor cells were divided evenly into a 384-well plate organized into a 24×16 grid (Fig. 6a) and stained with Hoechst 33342 solution. For drug screening, biologists added various FDA-approved drugs into the wells. Each drug was administered to a 1×8 column, starting with a $20\mu/mol$ dosage and tripling the dosage in each subsequent well (green box of Fig.6a). The last wells in the 1×8 column contained no drugs and were used as a control (blue box of Fig. 6a). Each well was imaged with the Operetta CLS™ high-content analysis system equipped with an high resolution 14bit CCD camera for cell imaging and the Harmony 3.5 software. Nine-field image montage per well (Fig. 6b) is generated by using an image stitching algorithm (Fig. 6c). The resolution of each image is 1360×1024 pixels. We took images from various locations with different drug dosages and evenly distributed cells, and made pairs for training set in which each pair consists of a bright-field image and its fluorescence nuclei image.

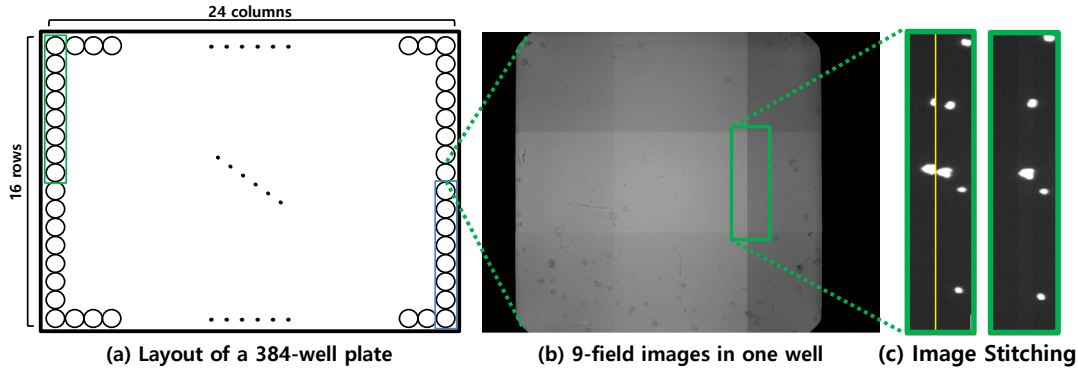


Figure 6: Overview of data acquisition and preprocessing: (a) layout of a 384-well plate, (b) nine overlapped images for a single well, (c) before and after stitching images (yellow line on the left is the border between adjacent images)

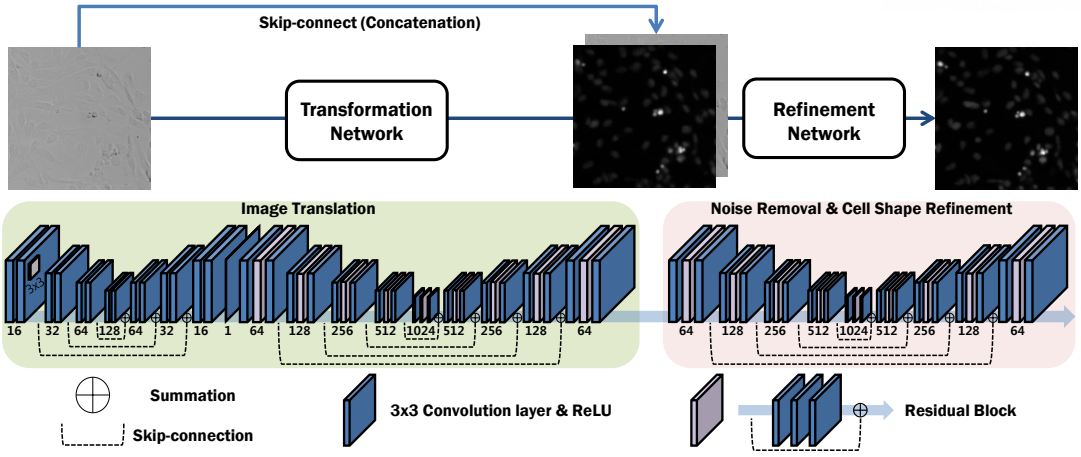


Figure 7: DeepHCS consists of two sub-networks: a Transformation Network (green box); and a Refinement Network (pink box). Convolution layers (blue layer) include ReLU as a non-linear function. Residual blocks (purple layer) consist of three identical convolution layers. All filter sizes used in this system are 3×3.

4.3 Proposed Method: DeepHCS

DeepHCS is built upon two deep neural networks, *Transformation Network (TN)* and *Refinement Network (RN)* (see Fig. 7).

Transformation Network is the first part of DeepHCS, consisting of two sets of FusionNet variant networks [56]. The first network in the TN is used to gradually transform the input bright-field image into the intermediate feature image, and the second network in the TN is used to actually perform the translation into the DAPI fluorescence image. The first network in the TN can effectively expand the depth of the network when the size of the input is relatively small, and adequately performs drop-out in which 16 feature maps are merged into one feature map at the end. The second network in the TN has more neuron weights by using residual blocks and actually translates the input image into DAPI fluorescence images. The number of filters in the entire network is expressed under each layer in Fig. 7.

Refinement Network is the second part of DeepHCS and is designed to improve the image quality of the translated result from the TN in terms of the noise and the cell shape. In contrast to the TN, the RN takes a concatenation of the translated TN result and the input bright-field image of the TN, which provides a clue to rectify errors in the translated image generated by the TN. For example, as shown in Fig. 8, the RN can improve the cell shapes and restore falsely removed cells. Another benefit of using the concatenated input image is to help reducing the gradient-vanishing problem caused by the black background in the translated result by the TN.

Loss Function For the TN, the mean-square error (MSE) is used to define the loss function to measure the pixel-wise error between the ground truth and the output image of the TN, as

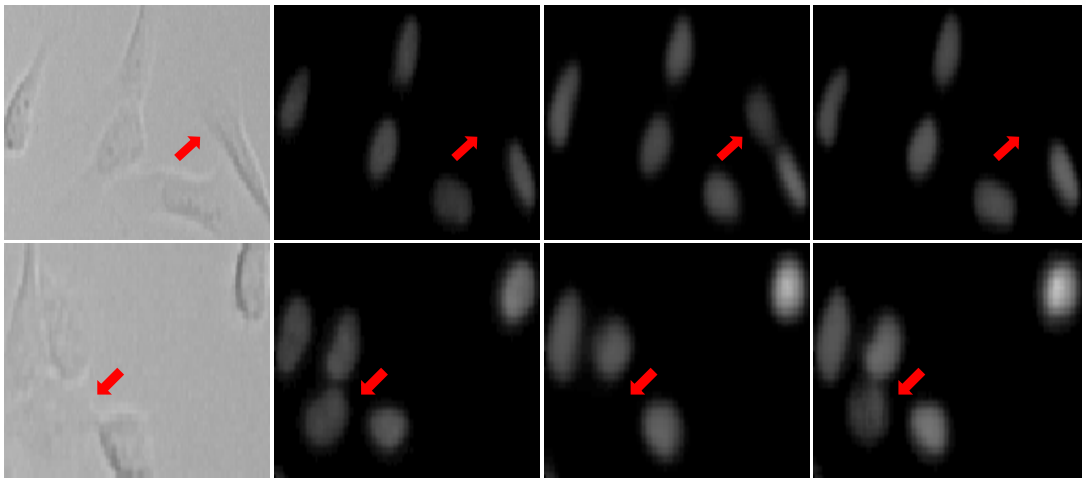


Figure 8: Refinement Network improves the cell shapes and restores missing cells. A false positive (top row) and a false negative (bottom row) from the TN are corrected by the RN.

follows:

$$L_{TN}(x) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{1}$$

where x is the input bright-field image, y_i is the real fluorescence image, and $\hat{y}_i$ is the output of the TN. For the RN, the mean-absolute error (MAE) and the SSIM are used as the loss function to deal with the shape of cells and the pixel intensity at the same time. The MAE is define as follows:

$$L_{MAE}(\hat{x}, y) = \frac{1}{n}\sum_{i=1}^{n}|r_i - y_i| \tag{2}$$

where $\hat{x}$ is the concatenation of the translated result of the TN and the input bright-field image, and r_i is the output of the RN. In contrast to the TN, we employed the MAE to handle the translated result of the TN because the MSE penalizes larger errors and is more tolerant to smaller errors [57]. The SSIM is defined as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

where μ_x and σ_x represent the mean and the variance of image x, respectively; σ_{xy} represents the covariance of image x and y, and c_1 and c_2 are two constant variables for division stability. Based on Eq. 3, we can measure the degree of structural change in the image and additionally recognize the difference between the two images based on luminance and contrast. The SSIM values range between 0 and 1; therefore, we defined the loss function using the SSIM as follows:

$$L_{SSIM}(\hat{x}, y) = \frac{1}{n}\sum_{i=1}^{n} 1 - SSIM(r_i, y_i) \tag{4}$$

By combining the two error measures, we can define the loss function for the RN as follows (α is empirically set to 0.8):

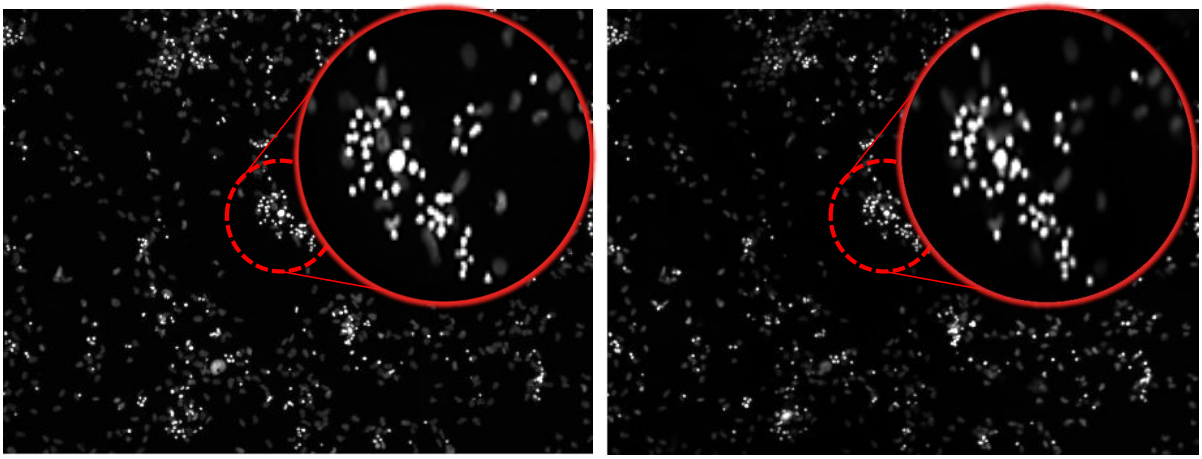
$$L_{RN}(\hat{x}) = (1 - \alpha) \cdot L_{MAE}(\hat{x}, y) + \alpha \cdot L_{SSIM}(\hat{x}, y) \tag{5}$$

4.4 Results

We used the training set consisting of 2,860 pairs of bright-field images and their corresponding fluorescence images, each measuring 256 × 256 pixels (we cropped the center of each image to reduce boundary effects). To validate DeepHCS, we used eight cases (C1 to C8), including either 1,386 or 2,646 images.

Table 1: Accuracy of the proposed method for eight test cases.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**
PSNR	33.91	33.90	33.79	33.93	38.52	39.04	38.65	38.46
SSIM	0.75	0.75	0.74	0.75	0.87	0.88	0.87	0.87
CVC	0.8663	0.9064	0.8794	0.8865	0.9583	0.9625	0.9673	0.9702



(a) Ground truth **(b) Ours**

Figure 9: (a) Ground truth fluorescence image, (b) the result of our method. Zoom-in shows the similarity of the cell shapes between the ground truth and ours.

To assess the quality of the images generated by DeepHCS, like the ablation study, we used two image error metrics (PSNR and SSIM) and cell viability correlation (CVC) that measures the similarity between the actual and generated DAPI fluorescence images using R^2 correlation, as shown in Table 1. In the experiment, we achieved an average of 0.9092 and a maximum of 0.9702 correlation with the ground truth. In addition, the shape of the cells and the status of the cells (living or dead) are clearly distinguished as shown in Fig. 9.

To further demonstrate the feasibility of DeepHCS for replacing biomarker generation in the conventional HCS workflow, we used seven other cases for the validation test. Fig. 10 shows the correlation of real DAPI images and our synthetic fluorescence images in terms of AUC and IC50, respectively. In addition, the responses of two anti-cancer drugs (AMG232 and RG7112) measured by AUC and IC50 are also shown using heatmaps; clear separation of two groups in drug responses are identically shown in DAPI images and ours. These results confirm that the

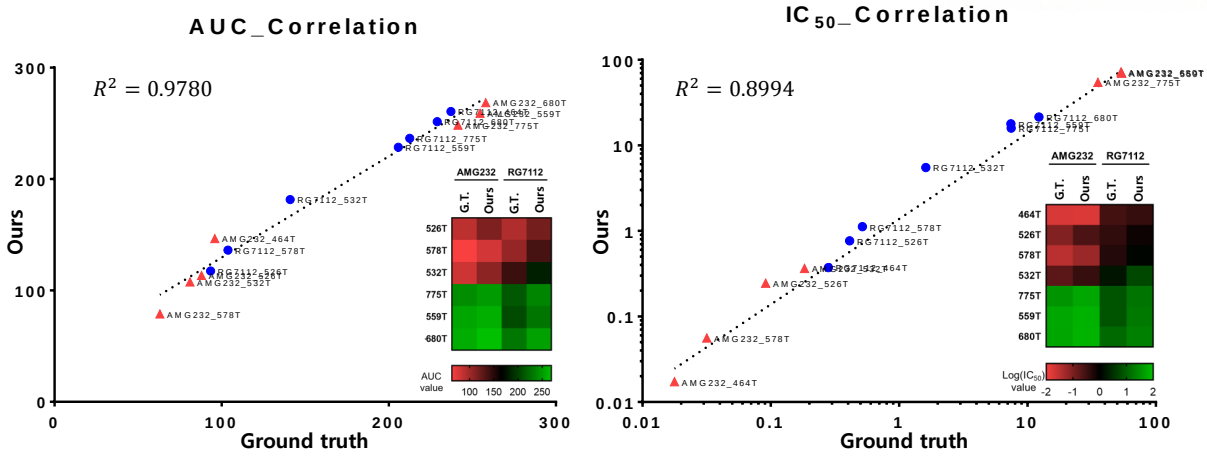


Figure 10: Comparison of AUC and IC50 values from the real DAPI images (ground truth) and our results from the seven patients' data. The heat maps show the drug response (green is low, and red is high).

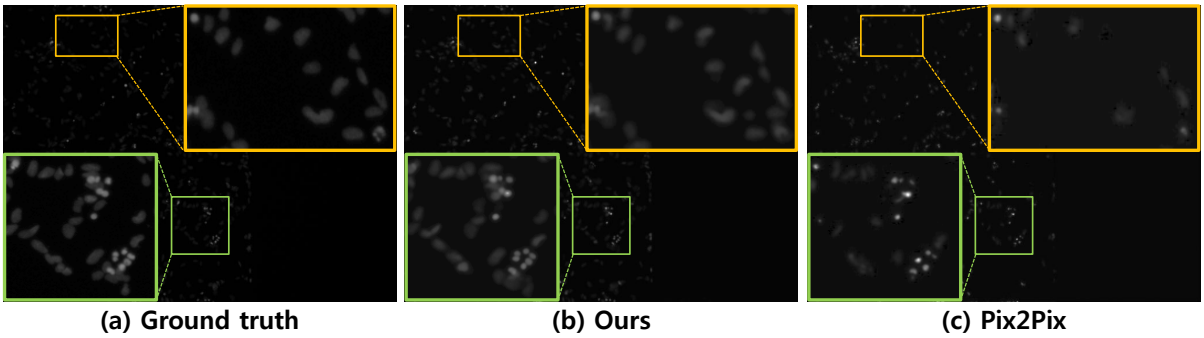


Figure 11: (a) Ground truth fluorescence image, (b) the results of the proposed method, (c) the results of the Pix2Pix network [1]. The results are generated after 300 training epochs. Our method can generate most of cell structures close to the ground truth.

images generated by DeepHCS can be used to compute AUC and IC50 for the estimation of drug responses, which shows potential to replace the conventional fluorescence imaging process in the HSC workflow.

Finally, we compared DeepHCS with the latest GAN-based image translation method used in the Pix2Pix network [1]. As shown in Fig. 11, DeepHCS reproduces cell images close to real DAPI images while Pix2Pix fails to accurately generate cell shapes in some cases. The GAN attempts to approximate the data distribution of the training set as much as possible; therefore, the reconstructed images look like cell images. However, this does not imply that the reconstructed image satisfies the accuracy up to the HCS analysis. Even though Pix2Pix learns the morphological characteristics in the cell image by minimizing the ℓ_1 distance from the ground truth, it is not enough to satisfy this problem.

4.5 Discussion

The effect of loss functions

Table 2: Different Losses on the TN and their accuracy for eight test cases. TN is a transformation networks. MSE, MAE, and SSIM represents the mean square error, mean absolute error, and structural similarity index, respectively.

Metrics	Nets+Loss	C1	C2	C3	C4	C5	C6	C7	C8
PSNR	TN_MSE	23.6277	22.9255	24.4305	24.3828	21.6432	21.2649	21.6310	21.1303
	TN_MAE	21.4519	20.6724	21.9138	21.9565	21.0067	20.7735	21.0504	20.7676
	TN_MSE_SSIM	21.4582	20.6527	21.9109	21.9635	21.0350	20.6608	21.0474	20.5306
	TN_MAE_SSIM	21.5188	20.7317	21.9950	22.0261	21.0675	20.7064	21.0836	20.5751
SSIM	TN_MSE	0.5879	0.5586	0.6194	0.6179	0.6429	0.6205	0.6449	0.6293
	TN_MAE	0.5187	0.4824	0.5317	0.5352	0.6507	0.6346	0.6516	0.6380
	TN_MSE_SSIM	0.5127	0.4743	0.5249	0.5288	0.6479	0.6301	0.6499	0.6363
	TN_MAE_SSIM	0.5101	0.4715	0.5229	0.5264	0.6457	0.6271	0.6479	0.6338

In order to minimize error values between real and generated images, several metrics can be nominated as a loss function according to their tasks. In our experiments, the MSE and MAE is widely used to minimize pixel-wise errors between real and generated images, and the SSIM is also used to focus on the degree of structural change. In Table 2, we observed that the difference of performance on the TN with combination of loss functions; MSE, MAE, MSE+SSIM, and MAE+SSIM. In both cases, PSNR and SSIM, the MSE is the best performance for the TN than other combination of loss functions. Based on this observation, it was determined that the MSE is most appropriate to quickly generate reference images on the basis of the RN process since the TN is first network to proceed with image translation in input images. For the RN, MAE and SSIM was considered as a loss function to revise the shape of cells so that the RN focus on structural difference between real and generated images based on luminance and contrast (SSIM) with pixel-wise errors (MAE) together.

Limitations

DeepHCS can transform bright-field images to DAPI fluorescence images, and then the drug response such as the cell viability, AUC, and IC50 in the translated DAPI fluorescence images is up to 97%, showing similar analysis results to real results. However, there are still two problems to be overcome. First, not only DAPI fluorescence images but also different modalities fluorescence images are actively used in HCS, but DeepHCS has the disadvantage in which the proposed method have to train different modalities fluorescence images. In order to overcome this problem, further research is needed to generate various types of fluorescence images capable of analyzing drug response and individual cells analysis over bright-field images.

Second, time-lapse drug response and image analysis are possible through removal of cell

fixation and staining, but this has not been verified due to the limitations of applicable time-lapse dataset. In subsequent research, relevant dataset is obtained to verify the possibility of analyzing drug response over time, which is not easily accessible in the conventional HCS pipeline.

4.6 Summary

In this contribution, we introduced DeepHCS, a novel deep end-to-end convolution neural network for generating DAPI fluorescence images directly from bright-field images. We showed that the DeepHCS can generate results similar to real DAPI images and outperforms state-of-the-art image translation methods. The proposed method demonstrates the potential to reduce the laborious biomarker preparation process and to improve the throughput of the large-scale image-based drug screening process using deep learning. In the future, we plan to apply the proposed method to time-lapse bright-field images, and assess the efficacy of generating other biomarker images.

V Bright-Field to Multiple Fluorescence Microscopy Image Conversion for Individual Cell Analysis in Label-Free High-Content Screening

5.1 Introduction

The effectiveness of high-content screening (HCS) in precision medicine has been demonstrated in recent studies [11, 55]. This approach to precision medicine involves readouts of various drug responses to patient-derived cell cultures, and uses high-throughput imaging and automatic image analysis to evaluate changes in the phenotype of the whole cells. The analysis is performed by counting the number of living cells versus dead cells, measuring the size of the cells, comparing the shape of the cells, and so on. In HCS, multiple imaging modalities are commonly used together to image various aspects of the cell phenotypes (Fig. 12). Such imaging modalities include bright-field and fluorescence microscopy images, in which the former captures the overall morphology of the cells, whereas the latter images various fluorescent biomarkers. One advantage of bright-field images is its ability to acquire a photographic record of cells without any extra preparation; fluorescence images require laborious cell fixation and staining procedures. Another advantage of the bright-field image method is its ability to capture the dynamics of cells during a period of time; fluorescence images can only generate a snapshot of cells at any given time (Fig. 12). Most of the previous work on bright-field microscopy image processing [28–30, 33] focused on cell segmentation and detection; however, the standard HCS workflow still heavily relies on detecting and analyzing biomarkers presented in fluorescence images.

With the advent of recent advances in deep learning, realistic image-to-image translation has become feasible [1,35,37]. Inspired by this research trend, in-silico labeling [26,52,54] has gained much attention in the biomedical field recently. The main benefit of in-silico labeling is the replacement of the laborious tissue preparation process with computational methods. However, the potential of in-silico methods have not been demonstrated in HCS applications. DeepHCS [2], introduced concurrently with the other in-silico labeling work, was the *first* in-silico labeling method that specifically targeted HCS for precision medicine. In this paper, we introduce *DeepHCS*$^{++}$, a novel data-driven image conversion technique for HCS to address several shortcomings in the previous work [2], such as limited ability to produce multiple fluorescence labels (only DAPI was tested) and lack of proper validation on live (time-lapse) cell images. We employed multi-task learning to concurrently generate multiple biomarkers and adversarial losses [13], which allowed for the generation of more realistic fluorescence microscopy images. We also tested the proposed method on three time-lapse data (each consisted of 65 time-lapse bright-field images). DeepHCS^{++} effectively avoided the time-consuming and laborious tissue preparation process for generating biomarkers while providing accurate image analysis results using the conventional image-based HCS workflow (Fig. 12 bottom row). The contributions of DeepHCS^{++} can be summarized as follows:

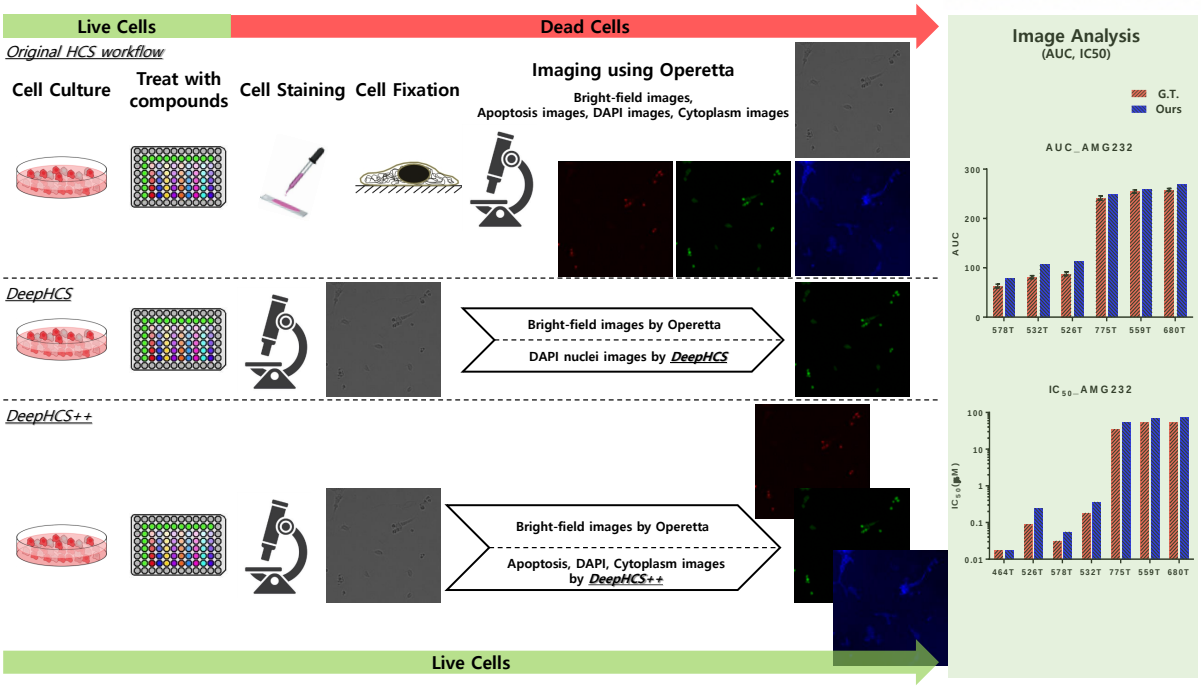


Figure 12: DeepHCS^{++} can eliminate the need for certain fluorescent markers hence simplifying the original HCS workflow resulting in skipping the process such as the cell fixation and staining. In contrast to DeepHCS [2], DeepHCS^{++} generates three different corresponding fluorescence images based on the bright-field image captured by Operetta. DeepHCS^{++} can keep the cells alive during the entire screening process; thus, it can follow a variant through several time-lapse images allowing us to observe the migration of cells over time.

- Using multi-task learning, the abundant and useful information obtained from additional sources based on three correlated tasks was generated in feature maps and the performance was better than in previous work.
- Employing adversarial loss helps the proposed method to generate more realistic fluorescence images.
- Our software-based workflow allows for observation of the drug response on living cells over time because staining and fixation are not required.

We evaluated the accuracy of DeepHCS^{++} using widely applied image quality metrics (e.g., peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM)). In addition, we compared our results with real apoptosis and DAPI fluorescence images based on dead cell number correlation (CNC) and cell viability [27], which demonstrates that the proposed method can replace the tissue preparation and fluorescence imaging processes in the conventional HCS workflow with a software-based image conversion process. Our results and real cytoplasm fluorescence images were also qualitatively validated using error map visualization and R^2 correlation.

5.2 Data

The data used in this study are imaged from glioblastoma (GBM) patient-driven cells and cell lines of GBM and lung adenocarcinoma. In this section, we explain the general data preparation and imaging process of the GBM data under the HCS setting because the same procedure is applied to the other data.

Tumor cells from GBM patients were cultured in a 384-well plate that was organized as a 24×16 grid of wells (Fig. 13a). Each well contained a nearly equivalent amount of tumor cells evenly spread across the entire well.

For drug screening, biologists medicated various FDA-approved drugs or targeted drugs into the wells. One drug was administered to a column of 1×8 wells (the green rectangle on the upper-left corner in Fig. 13a), with various dosage levels starting at $20\mu/mol$ multiplying by three for each succeeding well. One column of 1×8 wells (the blue rectangle on the lower-right corner in Fig. 13a) was used as a control bar (no drug was administered) to measure the relative drug response against other wells. Then, each well was imaged at $10\times$ magnification using an Operetta CLS™ high-content analysis system equipped with a high-resolution 14-bit charge-coupled device (CCD) camera for cell imaging; Harmony 3.5 software was used for analysis. The imaging system generated 4-channel output images, one channel for bright-field and the other three for fluorescence imaging. As shown in Fig. 13b, one well is covered by 9 images; the resolution of one image is 1360×1024 pixels, which corresponds to an actual physical dimension of 0.88×0.66 cm. Each image is then split into 256×256 pixel images. Only internal images in each well image were used for training; images containing black regions outside the well was discarded (Fig. 13b). We composed a training set from various locations with different drug dosages and evenly distributed cells. Each pair of training set consisted of bright-field images and their three corresponding fluorescence images.

In Fig. 4, an example of bright-field and the corresponding three fluorescence images are shown. The bright-field image (Fig. 4a) was created using the Operetta CLS™ microscopy system. Apoptosis fluorescence images (Fig. 4b) were stained using an Alexa 488 dye, which shows dead cells. DAPI fluorescence images (Fig. 4c) were stained using a Hoechst 33342 dye to detect the nuclei of cells (brighter cells are dead, the rest are alive). Cytoplasm fluorescence images (Fig. 4d) were stained using an Alexa 594 dye to observe the shape of cells. The top panel of the cytoplasm image in Fig. 4d does not have texture information because all the cells in this image are dead. Therefore, the size of the cells in this figure is smaller than normal. In general, the shape of the cells is stained, such as in the middle and bottom panels in Fig. 4d.

DeepHCS^{++} employs multi-task learning to concurrently generate three different fluorescence images with high image quality. DeepHCS^{++} is built upon two sequential deep convolutional neural networks, *transformation network (TN)* and *refinement network (RN)*; the discriminators are employed to improve the quality of the final result and make it more realistic in the RN (Fig. 14).

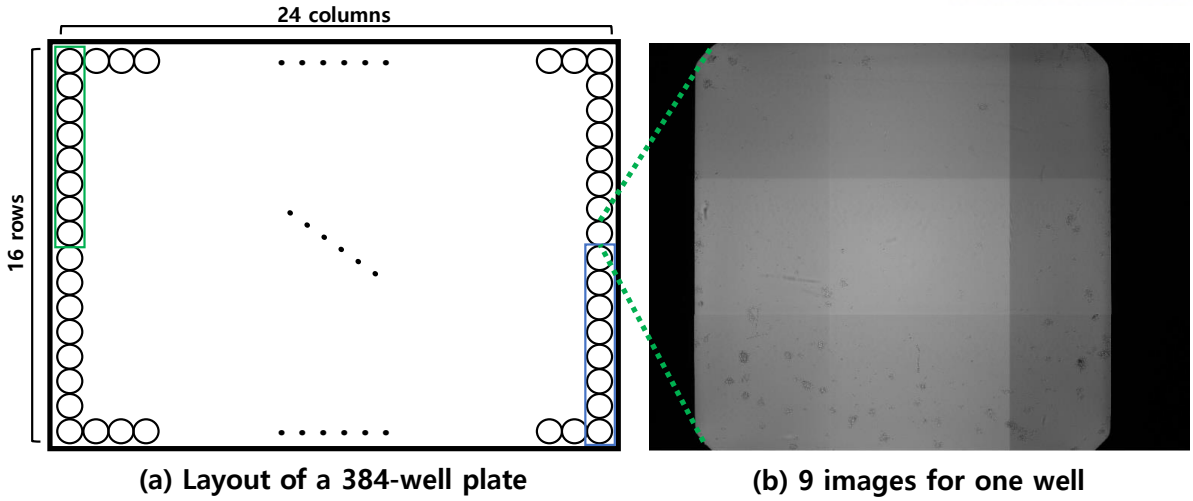


Figure 13: Overview of data acquisition: (a) layout of a 384-well plate, (b) nine overlapped images span a single well.

5.3 Proposed Method: DeepHCS^{++}

Architecture Overview

The proposed method involves two consecutive deep convolutional networks (Fig. 14). The transformation network (TN) translates a single bright-field to three fluorescence images of the apoptosis, nuclei, and shape of cells. The TN is built upon one encoder, which generates high-level compressed information. Three identical decoders take the compressed information and reconstruct corresponding fluorescence images. The second network, the refinement network (RN), is organized into three identical deep convolutional networks consisting of one encoder and one decoder each. The individual RN is allocated the bright-field image and each fluorescence image from the TN to revise and improve the image quality in terms of intensity, boundary, and shape of cells. The RN is also trained with adversarial loss so that the ability of the RN to create more realistic fluorescence images is improved. In the following sections, the TN and RN are explained in detail.

Transformation Network

Transformation network (TN) is the first part of DeepHCS^{++}, consisting of one encoder and three identical decoders for each of the three fluorescence images: apoptosis, DAPI, and cytoplasm. The first network in the TN is an encoder that takes a bright-field image as the input and generates high-level compressed information (1024 feature maps), which will be shared information for the three identical decoders. In the training process, the encoder translates the core information from the bright-field image into the final feature maps so that the decoders can conduct their corresponding task based on high-level compressed information such as the morphological shape, location of cells, and intensity of cells. The feature maps from the encoder

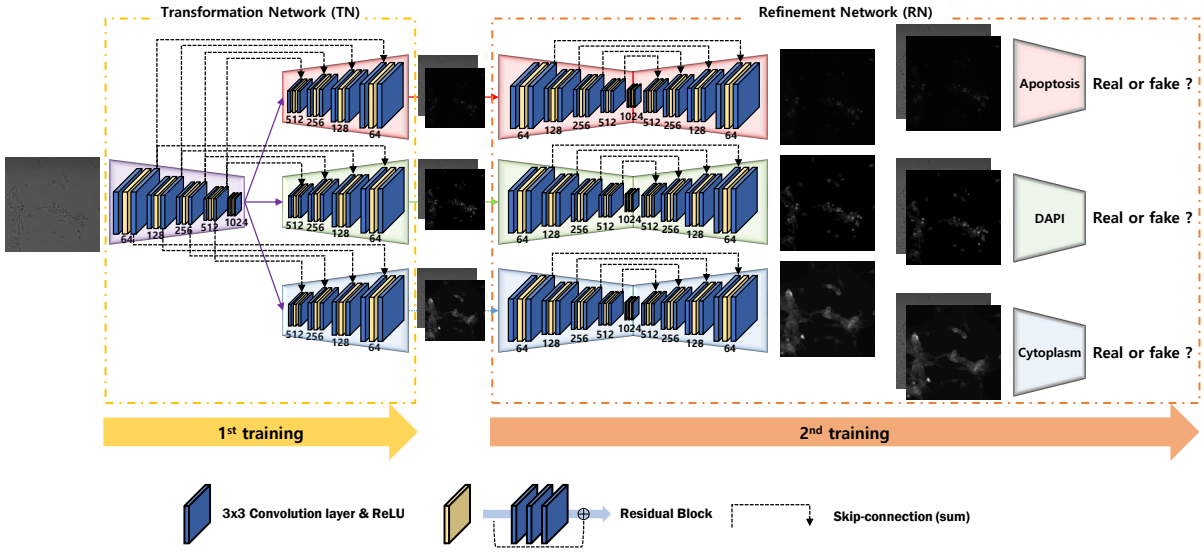


Figure 14: DeepHCS^{++} consists of two sub-networks, a transformation network (one encoder and three decoders); and a refinement network (three identical encoder-decoder networks). Convolution layers (blue) include ReLU as a non-linear function. Residual blocks (yellow) consist of three identical convolution layers. The filter size of all convolution layers used in the proposed method is 3×3. Max-pooling layers and upscale (nearest-neighbor interpolation) layers are adapted to scale down and up after a set of "convolution-residual block-convolution" layers.

contain abundant information and are essential for the ability of the decoders to transform shared features into each of the fluorescence images. As mentioned previously in earlier studies [49, 50], multi-task learning enables the network to obtain a better representation of feature maps through multiple tasks; thus, DeepHCS^{++} can produce various information in feature maps during the training process (this mention will be explained in Discussion section in detail).

The second network in the TN is composed of three identical decoders, which generate three fluorescence images for apoptosis, DAPI, and cytoplasm fluorescence images. With the exception of the last level of the encoder, the decoders and encoder have identical architecture. Skip-connections are connected between encoder and decoders via the same network level to help each level of decoders recover upper hierarchical information.

Refinement Network

Refinement network (RN) is the second network of DeepHCS^{++} and a variant of the networks of FusionNet [56]. Three identical networks revise three different fluorescence images from the TN. These networks are made up of the same architecture as TN, but each individual network is independently defined and trained for its task on corresponding fluorescence images. The RN is designed to improve the image quality of the translated images from the TN with respect to the intensity, boundary, and shape of cells. Each RN network takes a combination of the translated images from the TN and the bright-field image, which provides a clue on how to rectify errors

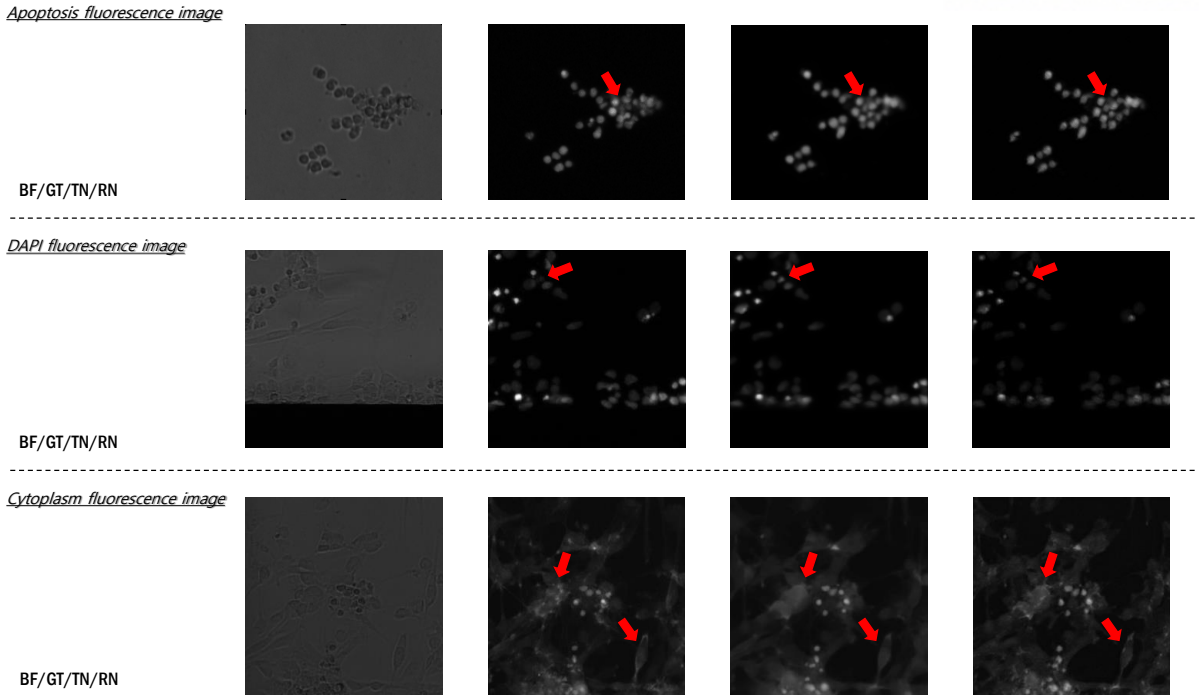


Figure 15: Each row represents bright-field (BF), ground-truth (GT) fluorescence image, the result from the TN, and the result from RN (left to right). Refinement network revises the translated images from the TN in terms of the intensity, boundary, and shape of cells. Red arrows indicate regions re-touched by the RN.

in the images translated by the TN. For example, as shown in Fig. 15, the RN can more clearly designate the boundary of cells (Fig. 15 upper panel) in the apoptosis fluorescence image. In the case of DAPI fluorescence images, dead cells are recovered by adjusting the intensity of cells (Fig. 15 middle panel). In the case of the cytoplasm fluorescence image, the translated image from the TN contains a blurry region; however, the blurry region is rendered more clearly by the RN (Fig. 15 bottom panel). Like the TN, skip-connections are also employed to conduct upper hierarchical information from the encoder to the decoder via a same-level network. Adversarial loss is utilized to increase the RN's ability to revise the images. A discriminator is defined for each of the three identical networks, where each takes a pair of bright-field and revised images from the RN, in addition to a pair of bright-field and ground-truth images, as their input. This allows the networks to distinguish whether each pair of bright-field and fluorescence images is real or fake.

Loss Function

For the loss function of the TN, the mean-square error (MSE) and the SSIM were used to measure the pixel-wise error between the real image and the results of the TN, as follows:

$$L_{MSE}(x) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{6}$$

where x is the input bright-field image, y_i is the real fluorescence image, and $\hat{y}_i$ is the result of each TN in terms of generated apoptosis, DAPI, and cytoplasm images.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7}$$

where μ_x and σ_x represent the mean and the variance of image x, respectively, σ_{xy} represents the covariance of image x and y, and c_1 and c_2 are two constant variables for division stability. Based on Eq. 7, we can measure the degree of structural change in the image and recognize the difference between the two images based on luminance and contrast. The SSIM values range between 0 and 1; therefore, we defined the loss function using the SSIM as follows:

$$L_{SSIM}(\hat{x}, y) = \frac{1}{n}\sum_{i=1}^{n}(1 - SSIM(r_i, y_i)) \tag{8}$$

We define the loss function for each TN by combining the two error measurements, as follows (α is empirically set to 0.8):

$$L_{TN_{apopt}}(\hat{x}) = (1 - \alpha) \cdot L_{MSE}(\hat{x}, y_{apopt}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{apopt}) \tag{9}$$

$$L_{TN_{dapi}}(\hat{x}) = (1 - \alpha) \cdot L_{MSE}(\hat{x}, y_{dapi}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{dapi}) \tag{10}$$

$$L_{TN_{cytop}}(\hat{x}) = (1 - \alpha) \cdot L_{MSE}(\hat{x}, y_{cytop}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{cytop}) \tag{11}$$

Total loss function for the TN is defined as follows:

$$L_{TN}(\hat{x}) = L_{TN_{apopt}} + L_{TN_{DAPI}} + L_{TN_{cytop}} \tag{12}$$

For the RN, the mean-absolute error (MAE) and the SSIM are used as the loss function to deal with the shape of cells and the pixel intensity at the same time. The MAE is defined as follows:

$$L_{MAE}(\hat{x}, y) = \frac{1}{n}\sum_{i=1}^{n}|r_i - y_i| \tag{13}$$

where $\hat{x}$ is the concatenation of the translated result of the TN and the input bright-field image, and r_i is the output of the RN. In contrast to the TN, we employed the MAE to handle the translated result of the TN because the MSE penalizes larger errors and is more tolerant of smaller errors. In the same manner as with TN, the SSIM values range between 0 and 1; therefore, we defined the loss function using the SSIM as follows:

$$L_{SSIM}(\hat{x}, y) = \frac{1}{n}\sum_{i=1}^{n}(1 - SSIM(r_i, y_i)) \tag{14}$$

By combining the two error measurements, we can define the loss function for the RN as follows (α is empirically set to 0.8):

$$L_{RN_{apopt}}(\hat{x}) = (1 - \alpha) \cdot L_{MAE}(\hat{x}, y_{apopt}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{apopt}) \tag{15}$$

$$L_{RN_{dapi}}(\hat{x}) = (1 - \alpha) \cdot L_{MAE}(\hat{x}, y_{dapi}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{dapi}) \tag{16}$$

$$L_{RN_{cytop}}(\hat{x}) = (1-\alpha)\cdot L_{MAE}(\hat{x}, y_{cytop}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{cytop}) \tag{17}$$

Furthermore, GAN [13] is an attractive concept and has been successfully used in many fields of research, such as computer vision. Pix2Pix [1], which also utilizes GAN, proposed the conditional generative adversarial network (CGAN) and achieved a good performance for image-to-image translation. Inspired by Pix2Pix, we used CGAN to improve the quality of results from the RN. Like Pix2Pix, each of the three RN networks becomes the generator, G; the three independent discriminators, D, are employed for generating apoptosis, DAPI, and cytoplasm fluorescence images. The conditional adversarial loss was employed for RN to improve the quality of its results. Conditional GANs is defined as follows:

$$L_{CGAN}(G, D) = \min_{G}\max_{D} \mathbb{E}_{x,y}[log(D(x, y))] + \mathbb{E}_{x,z}[log(1 - D(x, G(x, z)))] \tag{18}$$

where x is real input images, y is real label images, and z is generated images of the RN. G, the RN, tries to minimize, whereas D tries to maximize Eq. 18 that is separated according to generator and discriminator, respectively.

$$G_{CGAN} = \min_{G} \mathbb{E}_{x,z}[log(1 - D(x, G(x, z)))] \tag{19}$$

$$D_{CGAN} = \max_{D} \mathbb{E}_{x,y}[log(D(x, y))] + \mathbb{E}_{x,z}[log(1 - D(x, G(x, z)))] \tag{20}$$

The final loss functions for each RN are defined as follows:

$$L_{RN_{apopt}}(\hat{x}) = (1-\alpha)\cdot L_{MAE}(\hat{x}, y_{apopt}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{apopt}) + 0.1 \cdot G_{CGAN_{apopt}} \tag{21}$$

$$L_{RN_{dapi}}(\hat{x}) = (1-\alpha)\cdot L_{MAE}(\hat{x}, y_{dapi}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{dapi}) + 0.1 \cdot G_{CGAN_{dapi}} \tag{22}$$

$$L_{RN_{cytop}}(\hat{x}) = (1-\alpha)\cdot L_{MAE}(\hat{x}, y_{cytop}) + \alpha \cdot L_{SSIM}(\hat{x}, y_{cytop}) + 0.1 \cdot G_{CGAN_{cytop}} \tag{23}$$

Implementation details

Network architecture: DeepHCS^{++} consists of convolutional layers with non-linear activation functions to build up the TN and RN. We used a 3×3 convolutional layer with a stride of 1 and an ReLU function for the non-linear activation function. The zero-padding operation, “SAME" in tensorflow, was used to keep the spatial dimension of feature maps and the number of filters was doubled each level from 64 up to 1024. After a set of “convolution - residual block - convolution" layers, max-pooling or nearest-neighbor interpolation (up-sampling) was used when feature maps were down-sampled or up-sampled. At the end of the TN and RN, a 1×1 convolutional layer with a stride of 1 was used to predict the final translated fluorescence image.

Each discriminator in the RN consisted of five convolution layers where the number of filters is doubled in every layer from 64 to 512 except in the last (5th) convolution layer, which consists of only one 4×4 filter. The first three convolutional layers used a 4×4 filter size and a stride of 2 instead of using a pooling layer to down-sample feature maps. The last two convolution layers

used a 4×4 filter size and a stride of 1 to keep the spatial dimension of feature maps. After five convolution layers, a sigmoid function was used to produce the output of the loss function of the discriminator.

Training process: The proposed model was trained via two training steps. First, the TN was trained over 125 epochs and generated initial translated fluorescence images. Once the TN was fully trained, all of parameters of the TN are fixed. Based on generated fluorescence image and bright-field image, the RN is first trained over 20 epochs without discriminators to avoid early convergence of discriminator. Then, the RN is fully trained using conditional adversarial losses with MAE and SSIM losses over 100 epochs. During the training, we used the following two data augmentation methods: (1) randomly flipping the images horizontally or vertically, and (2) randomly rotating the images by $90°, 180°$, or $270°$. We used a mini-batch of size 4; the Adam optimizer was used with a learning rate of a 10^{-4}. The detail of training process is represented in Algorithm 1.

Algorithm 1: Training process of DeepHCS^{++}

Input: Set of input bright-field images X, Set of corresponding target fluorescence images Y

$N_{TN} \leftarrow$ the number of iteration for training TN

while $i \leq N_{TN}$ **do**

`/* step 1: Training the transformation network */`

$L_{TN_{apopt}}(\hat{x}) \leftarrow$*training TN for apoptosis by taking bright-field image* $\hat{x}$ *from* X

$L_{TN_{dapi}}(\hat{x}) \leftarrow$*training TN for DAPI by taking bright-field image* $\hat{x}$ *from* X

$L_{TN_{cytop}}(\hat{x}) \leftarrow$*training TN for cytoplasm by taking bright-field image* $\hat{x}$ *from* X

All of parameters of the TN are fixed

$N_{RN} \leftarrow$ *the number of iteration for training RN in early stage*

while $i \leq N_{RN}$ **do**

`/* step 2: Training the refinement network before adversarial learning */`

$L_{RN_{apopt}}(\hat{x}, y_{apopt}) \leftarrow$*training RN for apoptosis by taking* $\hat{x}$ *from* X *and generated fluorescence image from the TN*

$L_{RN_{dapi}}(\hat{x}, y_{dapi}) \leftarrow$*training RN for DAPI by taking* $\hat{x}$ *from* X *and generated fluorescence image from the TN*

$L_{RN_{cytop}}(\hat{x}, y_{cytop}) \leftarrow$*training RN for cytoplasm by taking* $\hat{x}$ *from* X *and generated fluorescence image from the TN*

$N_{Adv} \leftarrow$ *the number of iteration for training RN in adversarial stage*

α *is set to 0.1*

while $i \leq N_{Adv}$ **do**

`/* step 3: Adversarial learning with each RN */`

if $N_{Adv} == even$ **then**

$L_{RN_{apopt}}(\hat{x}, y_{apopt}) + \alpha \cdot G_{CGAN_{apopt}}$

$L_{RN_{dapi}}(\hat{x}, y_{dapi}) + \alpha \cdot G_{CGAN_{dapi}}$

$L_{RN_{cytop}}(\hat{x}, y_{cytop}) + \alpha \cdot G_{CGAN_{cytop}}$

if $N_{Adv} == odd$ **then**

$D_{CGAN_{apopt}} = \max_D \mathbb{E}_{x,y}[log(D(x,y))] + \mathbb{E}_{x,y}[log(1 - D(x, G(x, y_{apopt_{RN}})))]$

$D_{CGAN_{dapi}} = \max_D \mathbb{E}_{x,y}[log(D(x,y))] + \mathbb{E}_{x,y}[log(1 - D(x, G(x, y_{dapi_{RN}})))]$

$D_{CGAN_{cytop}} = \max_D \mathbb{E}_{x,y}[log(D(x,y))] + \mathbb{E}_{x,y}[log(1 - D(x, G(x, y_{cytop_{RN}})))]$

5.4 Results

We used the training set consisting of 8,064 pairs of bright-field images and their corresponding three fluorescence images from two completely different GBM patient-driven samples, in which each image is composed of 256×256 pixels as mentioned in Section 5.2.

The test dataset is also composed of GBM patient-driven samples from different patient samples.

Dataset1 was composed of two patient-driven samples. This dataset was treated with two groups of FDA-approved drugs (first group includes 21 drugs and second group includes 41 drugs). Each group was duplicated and tested for consistency to avoid a false response. In Table 3, 'Dup. 1 (Case 1)' and 'Dup. 2 (Case 2)' are the 21-drug groups and 'Dup. 1 (Case 3)', 'Dup. 2 (Case 4)' are the 41-drug groups for patient 1. Similarly, 'Dup. 1 (Case 5)' and 'Dup. 2 (Case 6)' are the 21-drug groups and 'Dup. 1 (Case 7)' and 'Dup. 2 (Case 8)' are the 41-drug groups for patient 2.

Dataset2 is the live cell image dataset that consists of the two cell lines of the GBM, the first type of cells is U87MG (Cell1) and the second type is L18 (Cell2). Each cell type was duplicated using two identical plates. One well plate of each cell type was treated with $50\mu/mol$ RG7112. The other well plate was left untreated. The live cell images were imaged over 16 hours with a snapshot taken every 15 minutes. Therefore, the case of Cell1 treated with RG7112 had 65 images (one starting image and 64 intermediate images). In summary, there are 4 cases in Dataset2: Cell1-RG7112, Cell1-NT (not treated), Cell2-RG7112, and Cell2-NT.

Dataset3 is the dataset of the A549 lung adenocarcinoma cell line, chosen to validate the generalization of our method over different cell types. Unlike the GBM case, the columns of 1×7 wells in one half of a 384-well plate are administered with TGF-β at various dosage levels starting from 0.005 ng/ml; then the amount was increased three times per well. In the other half of the 384-well plate, TGF-β is not administered in half remaining wells; whereas the galunisertib, an inhibitor, is administered to the last half at $10\mu/mol$. This dataset is stained with DAPI and cytoplasm to validate the applicability of DeepHCS^{++} on various cell types.

To validate the proposed method, we compared the generated fluorescence images with corresponding real fluorescence images by leveraging several metrics (PSNR, SSIM, CNC, CVC, R^2, and the changes in cell morphology) over Dataset1, Dataset2, and Dataset3.

Dataset1: patient-driven GBM cells

Result of apoptosis fluorescence image

To validate DeepHCS^{++} on the apoptosis fluorescence images, we used eight cases (Case 1 to Case 8 in Dataset1) where each case consists of either 1,386 or 2,646 images. As mentioned before, the apoptosis fluorescence image represents how many dead cells are left after drugs are administered. All dead cells from each image were counted to measure the CNC of dead cells in generated and real apoptosis fluorescence images in Fig. 16; their R^2 correlation values

Table 3: Accuracy of PSNR, SSIM, and cell number correlation (CNC) for eight test cases of apoptosis fluorescence images (Dataset1).

Method	Metric	Patient 1				Patient 2			
		Drug group 1		Drug group 2		Drug group 1		Drug group 2	
		Dup. 1	Dup. 2	Dup. 1	Dup. 2	Dup. 1	Dup. 2	Dup. 1	Dup. 2
DeepHCS++	PSNR	34.581	35.008	35.166	35.212	36.512	36.982	36.655	36.863
	SSIM	0.735	0.749	0.743	0.745	0.805	0.820	0.811	0.812
	CNC	0.9157	0.9265	0.9012	0.8987	0.9458	0.9467	0.9501	0.9487

Table 4: Accuracy of PSNR, SSIM, and cell viability correlation (CVC) for eight test cases of DAPI fluorescence images (Dataset1).

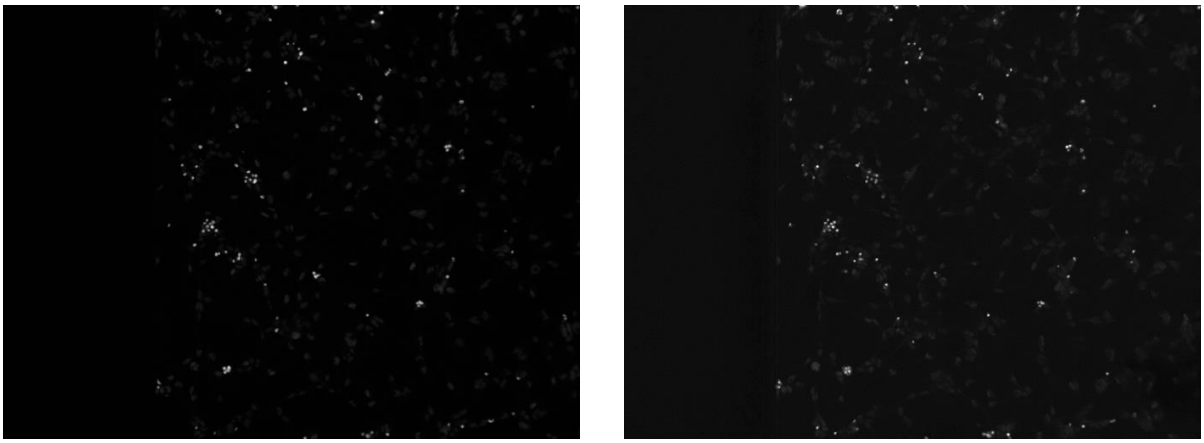
Method	Metric	Patient 1				Patient 2			
		Drug group 1		Drug group 2		Drug group 1		Drug group 2	
		Dup. 1	Dup. 2	Dup. 1	Dup. 2	Dup. 1	Dup. 2	Dup. 1	Dup. 2
DeepHCS++	PSNR	33.707	33.743	33.618	33.741	38.264	38.65	38.385	38.226
	SSIM	0.786	0.788	0.777	0.782	0.893	0.899	0.897	0.895
	CVC	0.9564	0.956	0.934	0.9348	0.973	0.9798	0.9785	0.9757

were computed to measure the similarity between the two results, as shown in Table 3. In this experiment, we achieved an average of 0.8987 and a maximum of 0.9501 correlation between the real and generated fluorescence images. PSNR and SSIM were also measured to validate the visual similarity between generated and real apoptosis fluorescence images. The highest PSNR and SSIM values were 36.982 and 0.820, respectively (Table 3). We can observe that PSNR, SSIM, and CNC are similar for the same drug group in the same patient-driven sample, which means that the duplicated drug group was highly consistent and avoided false responses.

Result of DAPI fluorescence images

Similar to the result of apoptosis fluorescence images, we used the same eight cases (Case 1 to Case 8 in the Dataset1) of either 1,386 or 2,646 images to validate the performance on the DAPI fluorescence images. To assess the quality of DAPI fluorescence images generated by DeepHCS^{++}, PSNR, SSIM, and cell viability correlation (CVC) were used to measure the similarity between the real and generated DAPI fluorescence images over the Dataset1, as shown in Table 4. The cell viability is a numerical value defined by the ratio between the number of live cells and the number of total cells in each well. The CVC is computed using R^2 correlation between real and generated cell viability in DAPI fluorescence images, and is a very important value when measuring the drug response. In this experiment, DeepHCS^{++} achieved an average of 0.934 and a maximum of 0.9798 correlation on the CVC. In addition, the shape of the cells and the status of the cells (living or dead) are clearly distinguished, as shown in Fig. 17.

Result of Cytoplasm fluorescence images



(a) Ground-truth (b) Ours

Figure 16: (a) Ground-truth fluorescence image, (b) result of our method in the apoptosis fluorescence image. These images from our method focus on translating with respect to dead cells.

The cytoplasm fluorescence image represents various cell shapes (Fig. 4d). The morphological features in each cell were analyzed for change in response to different drugs. Figure. 18 shows the per-pixel error map between real and generated cytoplasm fluorescence images where the error value ranges from -255 to 255. The error maps visualize the positive (cyan) and the negative (purple) differences. The black color regions in the error map represent the zero pixel value when the real and generated images are identical. We can see that the real and generated cytoplasm fluorescence images are very similar in Fig. 18 because the color of each pixel in the error map is close to black.

In order to measure the similarity of pixel intensity, R^2 correlation was also calculated, as shown in Fig. 19. Each row represents one field image randomly selected from Case 1 to Case 8 in Dataset1. The generated cytoplasm fluorescence images have entirely low pixel values compared to the real cytoplasm fluorescence image, therefore, R^2 correlation is relatively low even though the visualized quality of real and generated cytoplasm fluorescence images was good. Figure. 20 explains that the qualitative results of each image from DeepHCS^{++} are highly uniform because the gap between average R^2 correlation in each ease (Case 1 to Case 8 in Dataset1) is very small (0.7379 to 0.7558). The estimated figure confirms that DeepHCS^{++} can consistently generate cytoplasm fluorescence images in Fig. 20.

Dataset2: Live cell (time-lapse) image

In Section 5.1, we mentioned that DeepHCS^{++} can reduce the reliance on fluorescence protein markers, therefore, it reduces the level of induced phototoxicity and potentially cytotoxicity during the imaging process and cell staining, respectively. To assess the analysis of live cells, a cell line was newly cultured and imaged (Dataset2). In Dataset2, a total of four cases including

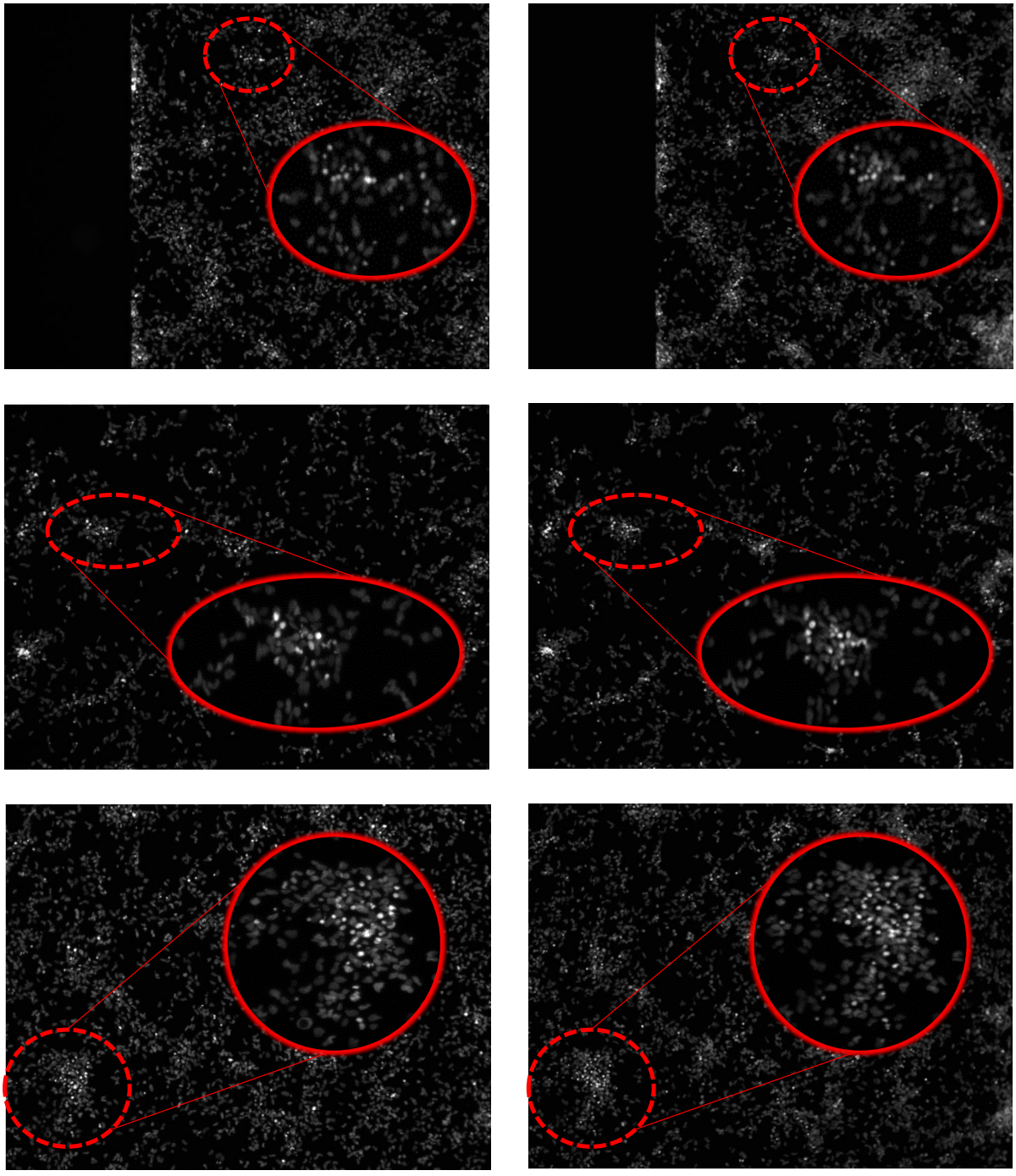


Figure 17: (a) Ground-truth fluorescence image, (b) DAPI fluorescence image. Close-up shows the similarity of the cell shapes between the ground truth and the result of our method.

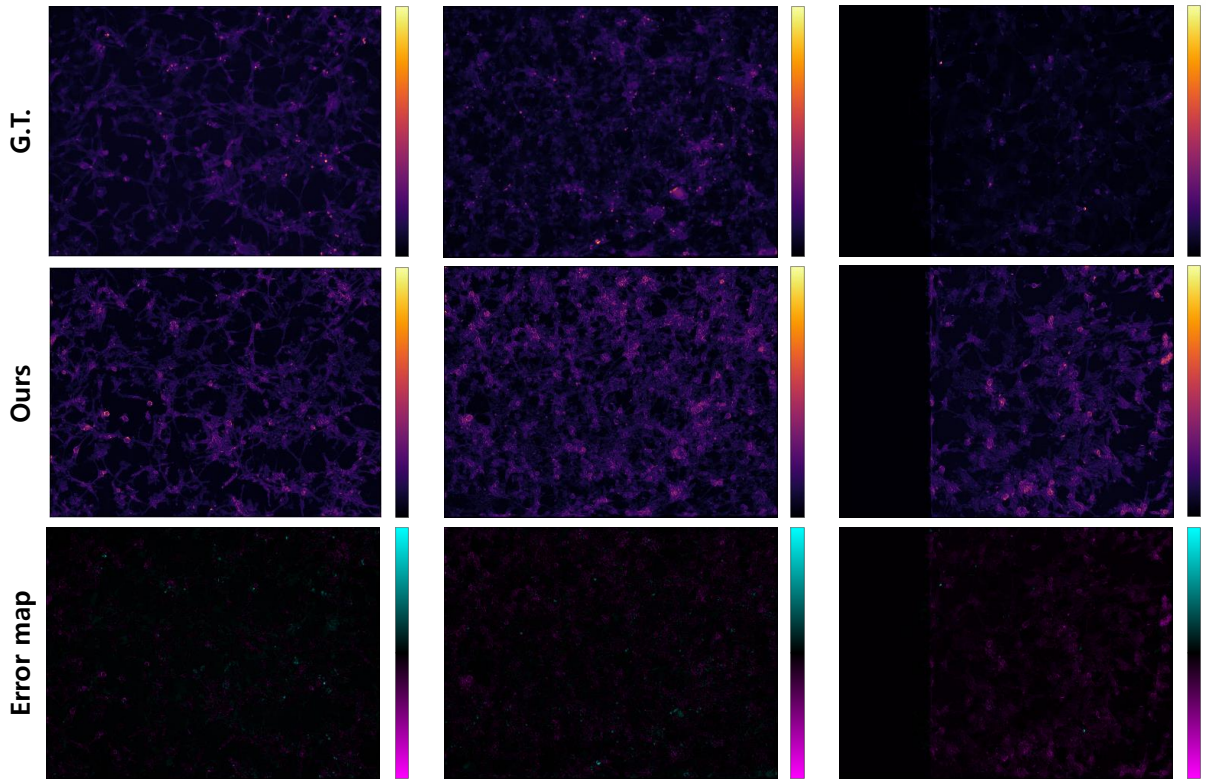


Figure 18: Error maps are visualized to identify the difference between generated and real cytoplasm fluorescence image ranging from -255 to 255.

U87MG with RG7112 (Cell1-RG7112), U87MG without RG7112 (Cell1-NT), LN18 with RG7112 (Cell2-RG7112), and LN18 without RG7112 (Cell2-NT) were used in the live cell experiment. Each case was imaged once every 15 minutes for 16 hours, resulting in 65 time-lapse bright-field images and their corresponding DAPI fluorescence images. In contrast to the original training process for DeepHCS^{++}, one case (Cell1-RG7112) of Dataset2 is fed into the pre-trained DeepHCS^{++} to perform the transfer learning because the cancer cell lines and patient samples have significantly different morphological features (shape, size, etc.) and imaging settings. In this case, the TN and RN were trained together with 100 epochs; the Adam optimizer was used with a learning rate of 10^{-4}.

To validate the performance of live cell images, the averaged migration of cells in each image was measured using Cell Tracker [58]. Cells in the generated and real DAPI images were detected (Fig. 21), and their movement (cell migration) over time was traced. In Fig. 22, we visualize the averaged cell migration per one time point in the generated and real DAPI fluorescence images. The trend of the cell migration over 64 points (16 hours) is similar in real and generated fluorescence images.

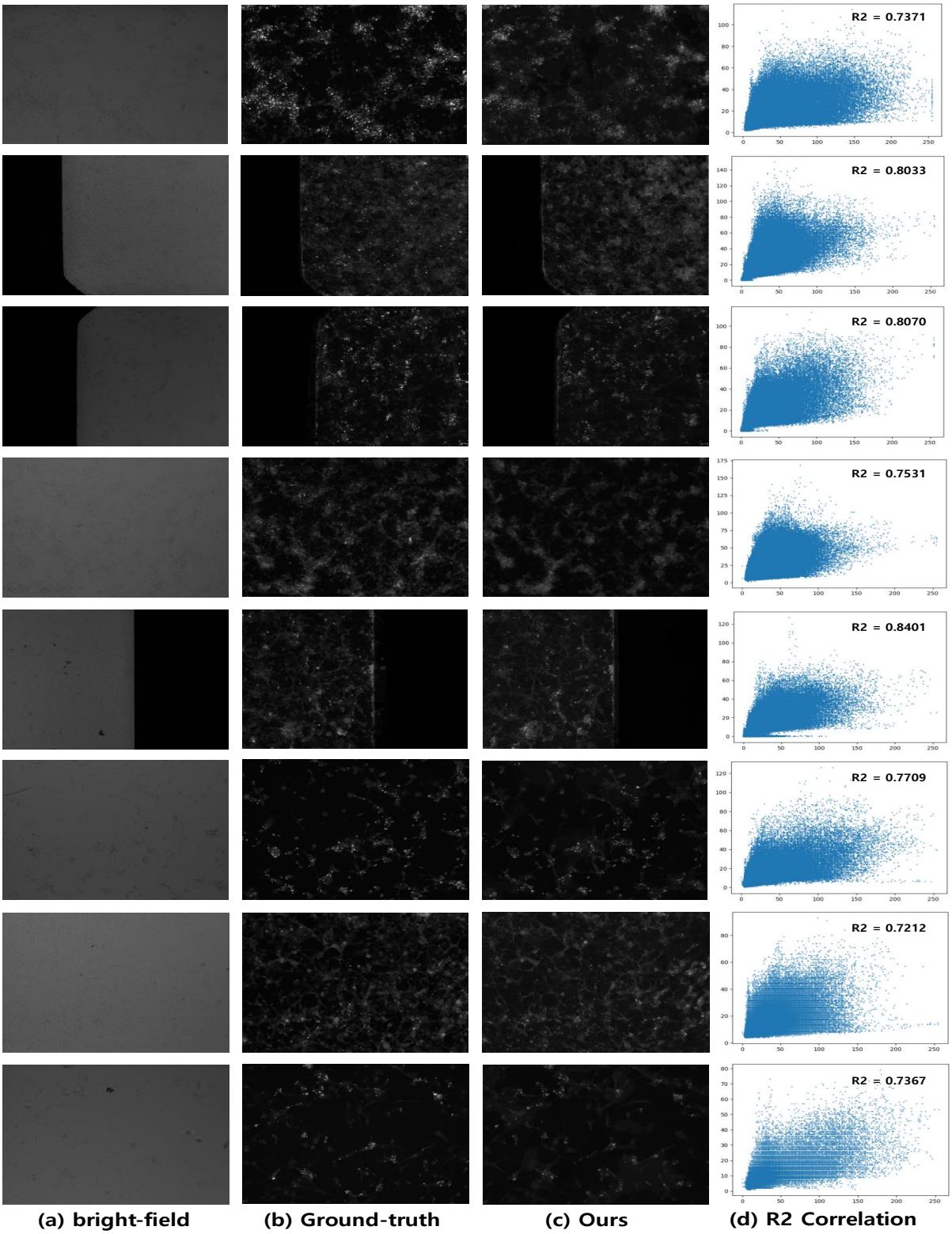


Figure 19: Each row represents one image randomly selected from Case 1 to Case 8 in Dataset1 (top to bottom): (a) bright-field image as input for the proposed method; (b) real cytoplasm fluorescence image; (c) generated cytoplasm fluorescence image by DeepHCS^{++}; (d) R^2 correlation between (b) and (c). The x-axis and y-axis of (d) represent the pixel value of the real cytoplasm fluorescence and generated cytoplasm fluorescence image, respectively.

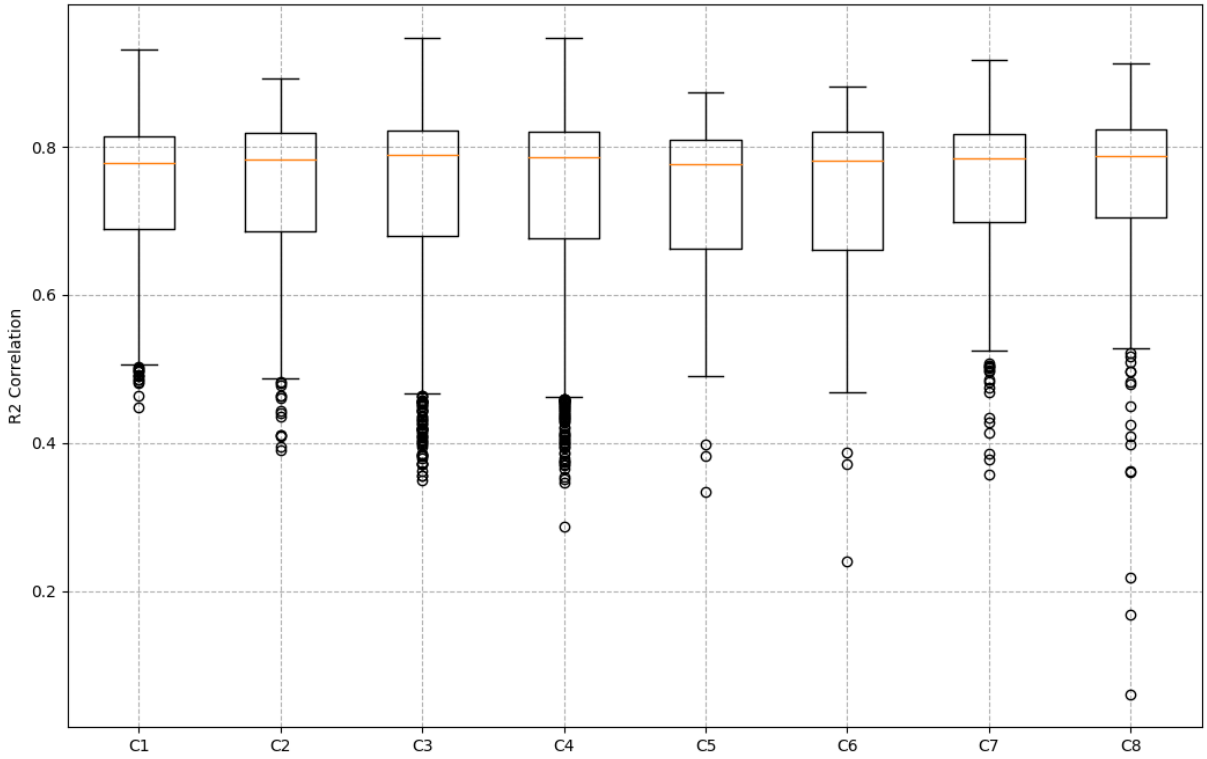


Figure 20: Boxplot of R^2 correlation of cytoplasm fluorescence images between real and generated images from C1 to C8. The average of each case (C1 to C8) is 0.7427, 0.7451, 0.7387, 0.7379, 0.7431, 0.7463, 0.7558, and 0.7573. The average R^2 correlation in the entire Dataset1 is 0.7459.

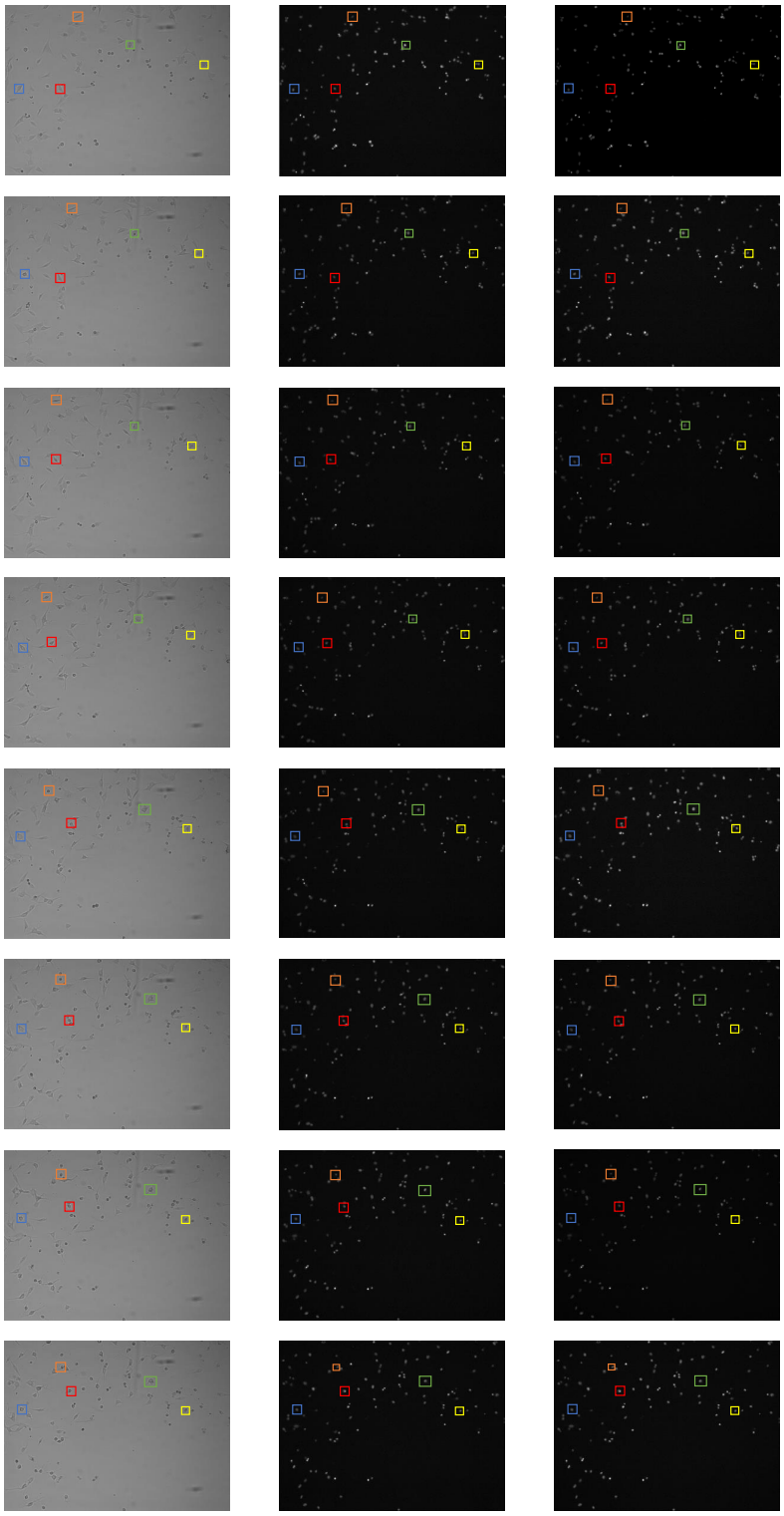


Figure 21: Each row represents different time-lapse images. 65 time points were used for live cell imaging. This figure shows the images at 1, 9, 17, 25, 33, 41, 49, and 57 time points, respectively (from top to bottom). Cells are tracked over time (rows) and across different images (columns) using colored boxes. The supplemental file contains generated time-lapse images as a movie to compare it with real time-lapse images.

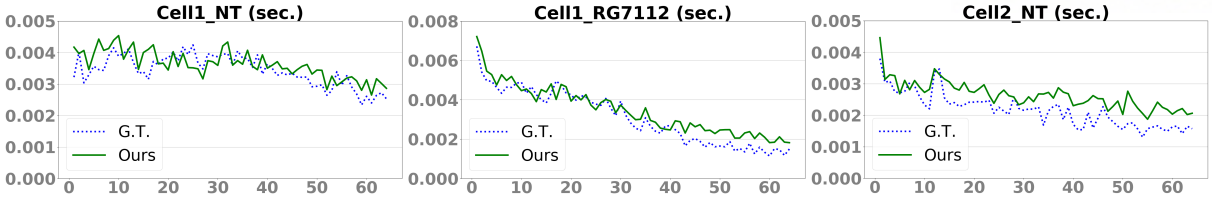


Figure 22: Migration of live cells between real and generated DAPI fluorescence images. The trend of the cell migration is similar between real and generated DAPI fluorescence images. The time unit is seconds.

Dataset3: lung adenocarcinoma cell line

In the previous study [2], only patient-driven GBM datasets were used to validate their method. To address this limitation, DeepHCS^{++} validated the generalization to another cell type, the A549 lung adenocarcinoma cell lines (Dataset3), by generating fluorescence images from bright-field images (Fig. 23). For Dataset3, DeepHCS^{++} was performed the transfer learning with pre-trained weights of the GBM-data and generated fluorescence images were measured, with a trend of cell morphological change between ground-truth and our method. Dataset3 is imaged with same device setting in Section 5.2; the changes in cell morphology caused by the transforming growth factor-β (TGF-β) were observed on cell length and width. The restrained growth caused by an inhibitor (galunisertib) was also observed.

It is well known that TGF-β induces epithelial-to-mesenchymal transition (EMT) which is a critical process during development and cancer progression [59]. Key features of EMT are cell morphological changes such as increased cell length and the spindle-like shape [60]. In this experiment, we generated fluorescence images (Fig. 23) of TGF-β-induced EMT in A549 lung adenocarcinoma cell lines and measured the average length, width, and width-to-length ratio of cells per well by performing segmentation of each cell using Cellpose [61] to validate a correlation between real and generated fluorescence images. The R^2 correlation of cell morphology is over 0.8, which means that the trend of cell morphological changes are similar to all cases with growth factor, growth factor+inhibitor, and no growth factor (Fig. 24). This result suggests that DeepHCS^{++} is not cell type-specific and can be generalized over various cell types other than GBM. In addition, DeepHCS^{++} successfully analyzes cell cytoplasm.

In summary, the images generated by DeepHCS^{++} can be used for live cell image analysis as well as image-based drug response analysis and cell analysis over various cell types, which shows the potential of replacing the conventional fluorescence imaging process with our method.

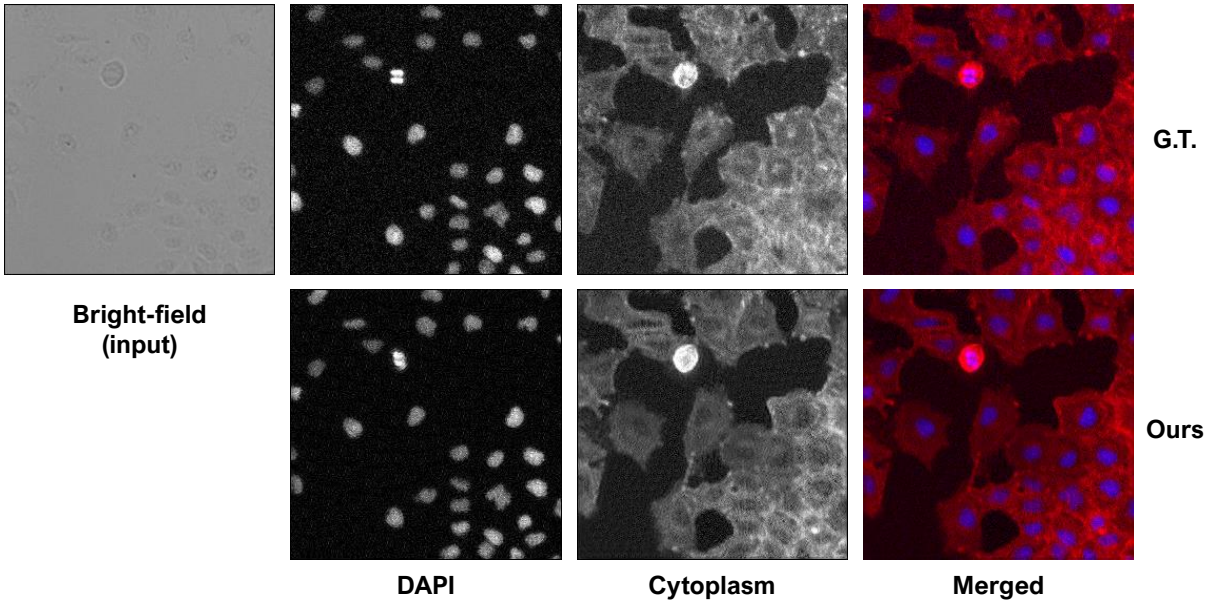


Figure 23: Visual quality of generated fluorescence images from DeepHCS^{++} is acceptable although the employed dataset is different from the GBM dataset. The last column shows the overlays of DAPI (blue) and cytoplasm (red) images.

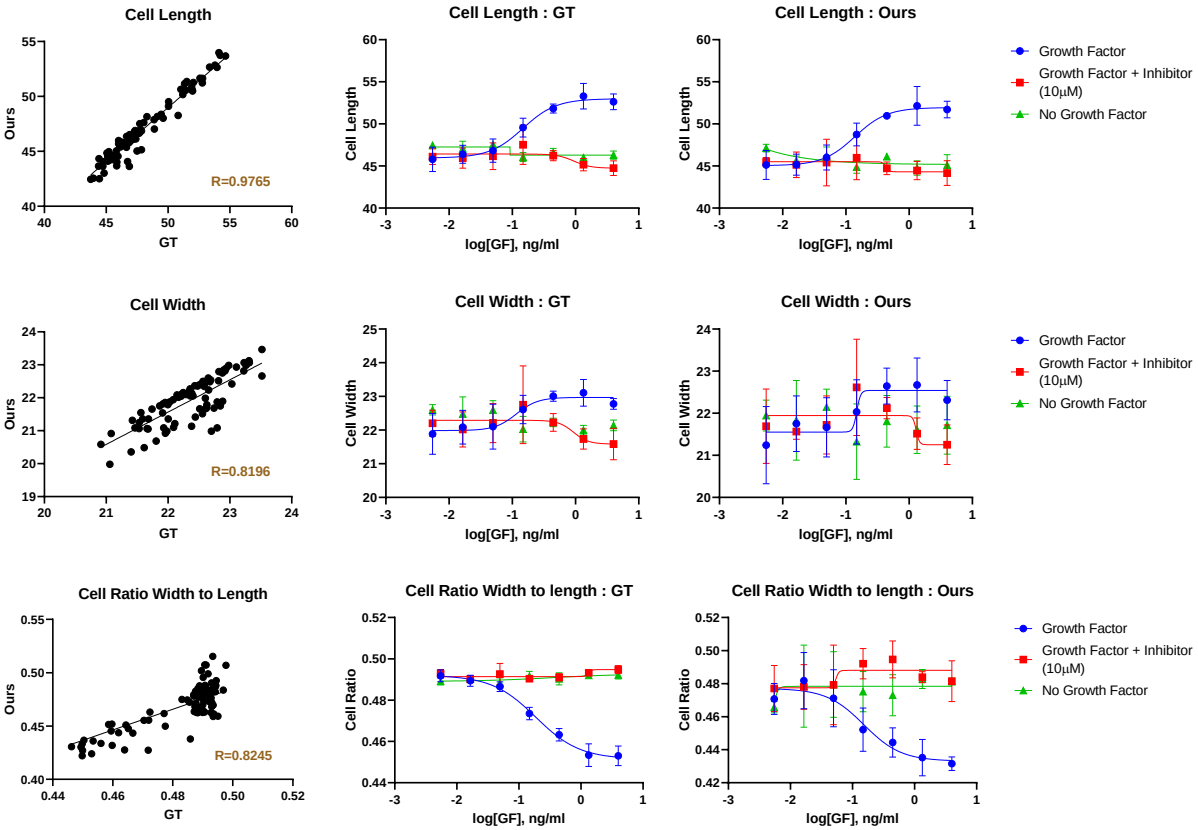


Figure 24: Correlation between the ground-truth and DeepHCS^{++}-generated images over the average length, width, and width-to-length ratio of cells per well. All the R^2 values are over 0.8 (first column), showing high correlation. The trends of cell morphological changes between two groups are also similar (second and third columns).

5.5 Discussion

DeepHCS^{++} extends DeepHCS [2] by addressing its limitation of generating only DAPI fluorescence images from bright-field images. In DeepHCS^{++}, we generalized the method to handle multiple biomarkers in a single multi-task learning network and improved the performance of the original method.

In this section, we compare DeepHCS^{++} with existing image translation methods (DeepHCS and Pix2Pix [1]), and discuss the strength and limitations of the proposed method.

Comparison with DeepHCS and Pix2Pix

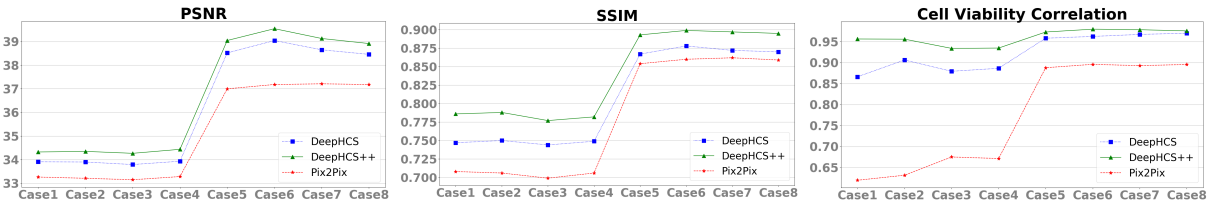


Figure 25: Comparison between DeepHCS^{++} and DeepHCS and Pix2Pix using PSNR, SSIM, and CVC. The x-axis of all graphs represents case 1 to case 8 in Dataset1. The y-axis represents their corresponding measured values.

In [2], the original DeepHCS method was compared with Pix2Pix, one of the commonly used conventional image translation methods; the advantage of the proposed method was demonstrated, especially in microscopy image translation. In this section, we will compare these two methods with DeepHCS^{++} on DAPI fluorescence images (see Fig. 25). We used three quality metrics, PSNR, SSIM, and CVC, to measure the pixel-level and structural (i.e., shape) similarities between generated and ground-truth biomarker images and the accuracy of detected live and dead cells, respectively. We observed that DeepHCS^{++} can achieve higher accuracy across three metrics over eight test cases compared to DeepHCS, which we believe is due to the multi-task and adversarial learning employed in our method (the effect of multi-task learning will be discussed in the following section).

Effect of Multi-task Learning

Multi-task learning can be classified into two types: the hard parameter sharing and soft parameter sharing [50]. The former shares the middle layers with the reconstruction layer at the end of each task; whereas, the latter defines independent networks for each task but all layers are constrained by the loss function enforcing the similarity between networks. DeepHCS^{++} belongs to the former case (hard parameter sharing) because the encoder in the TN is shared by three decoders for biomarker images. Ruder *et al.* [50] explained the advantages of multi-task learning based on several observations. The comprehensive viewpoint according to these observations argues that richer information is composited to hard parameter sharing layers to deal

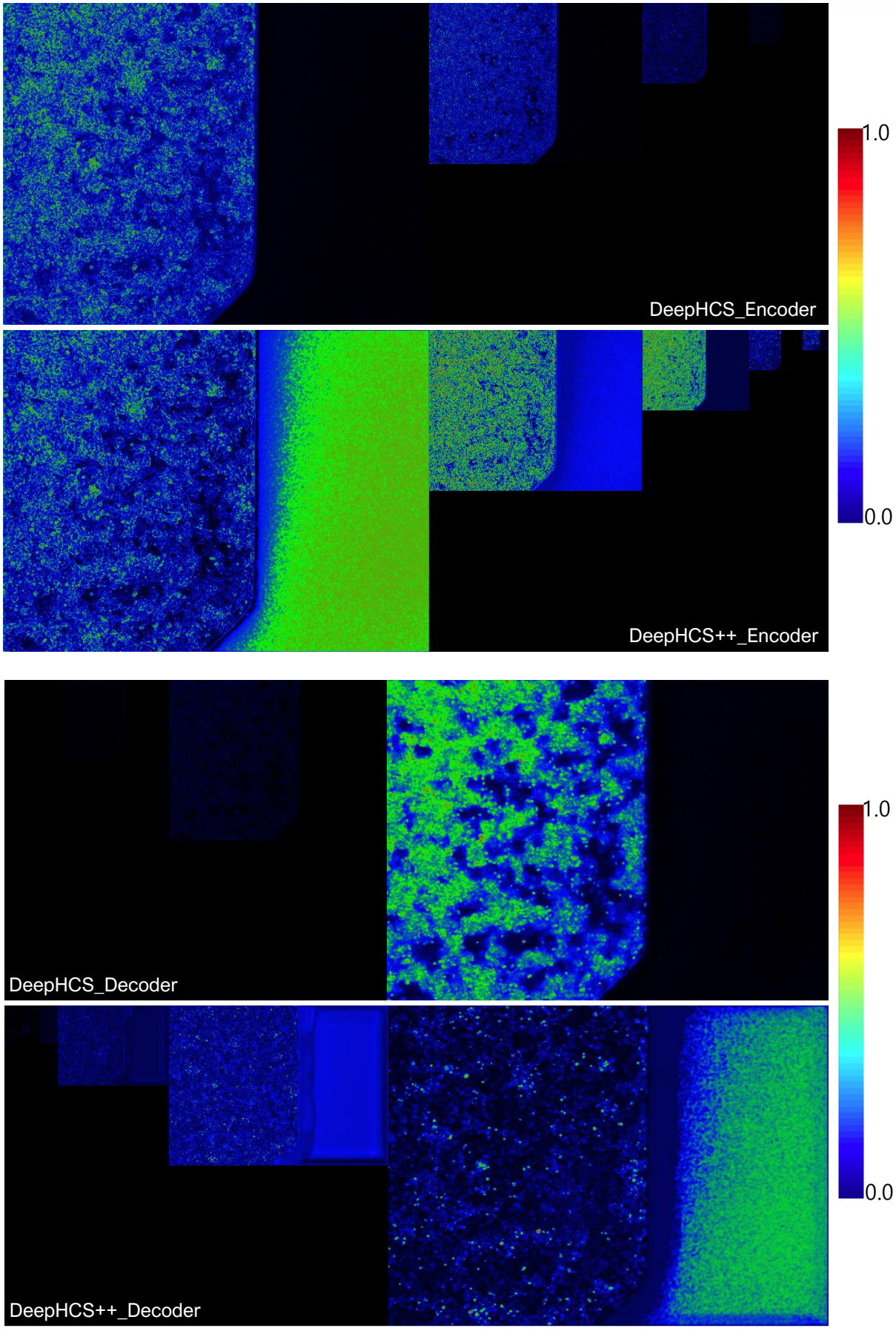


Figure 26: Difference between feature maps of DeepHCS and DeepHCS^{++}. Each feature map is extracted after the last convolution and activation layer in each level layer from the decoder part of TN and averaged along the channel axis.

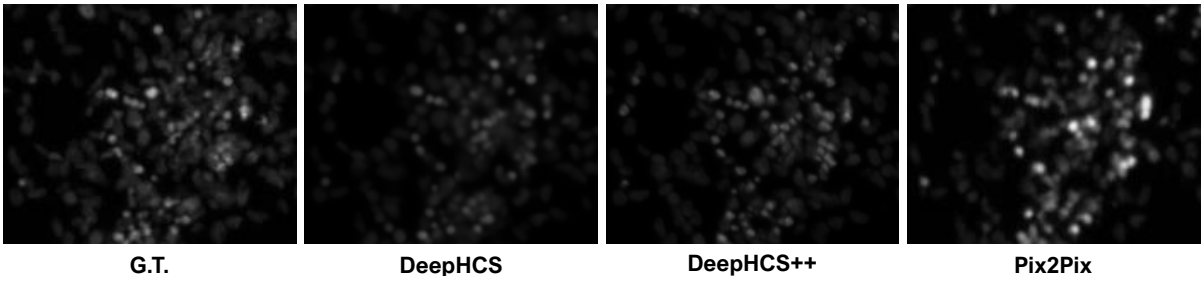


Figure 27: Close-up images showing the difference among the three methods. In case of Pix2Pix, live cells are not well reconstructed, whereas dead cells are relatively well reconstructed.

with multiple tasks together, which applies to our case as well. To validate this observation, we visualize and compare the feature maps of DeepHCS and DeepHCS^{++} in Fig. 26. Note that strong feature responses are evenly distributed in every level in DeepHCS^{++} whereas features in DeepHCS are relatively weak except in the very first and last level in the TN. This demonstrates that multi-task learning in DeepHCS^{++} helps to flow information across different levels, which contributes to storing more information in the network compared to the model based on single-task learning.

Adversarial Learning

In contrast to DeepHCS, DeepHCS^{++} employs adversarial learning in the RN to generate more realistic and high-quality fluorescence images. Figure. 27 shows that DeepHCS^{++} generates images with sharper and clearer cell boundaries, whereas DeepHCS generates blurrier images. We also made a comparison with Pix2Pix, which is a GAN-based image translation method, to assess the performance of DeepHCS^{++}. The same training dataset (8,064 DAPI fluorescence images) was used for training Pix2Pix, and it was deployed on Dataset1 for testing. Figure. 27 shows that Pix2Pix reconstructs dead cells relatively well, whereas live cells are very blurry; some of them are even missing. Even though Pix2Pix learns the morphological components of cells by minimizing the ℓ_1 distance from the ground-truth, this seems insufficient to encode the shape and texture of the biomarker images. In addition, the pixel intensity distribution of the DeepHCS^{++}-generated image is closer to that of the ground-truth (Fig. 28). We believe that the adversarial learning employed in our method helps match the distribution of pixel intensity in the training set, which eventually contributes to improve SSIM, CNC, and CVC as shown in Fig. 25.

Limitations

Even though DeepHCS^{++} shows potential for avoiding the time-consuming tissue preparation process and improving the throughput of image-based drug screening process, there still exist several limitations.

One limitation is that some biomarkers cannot be inferred only from bright-field images. One

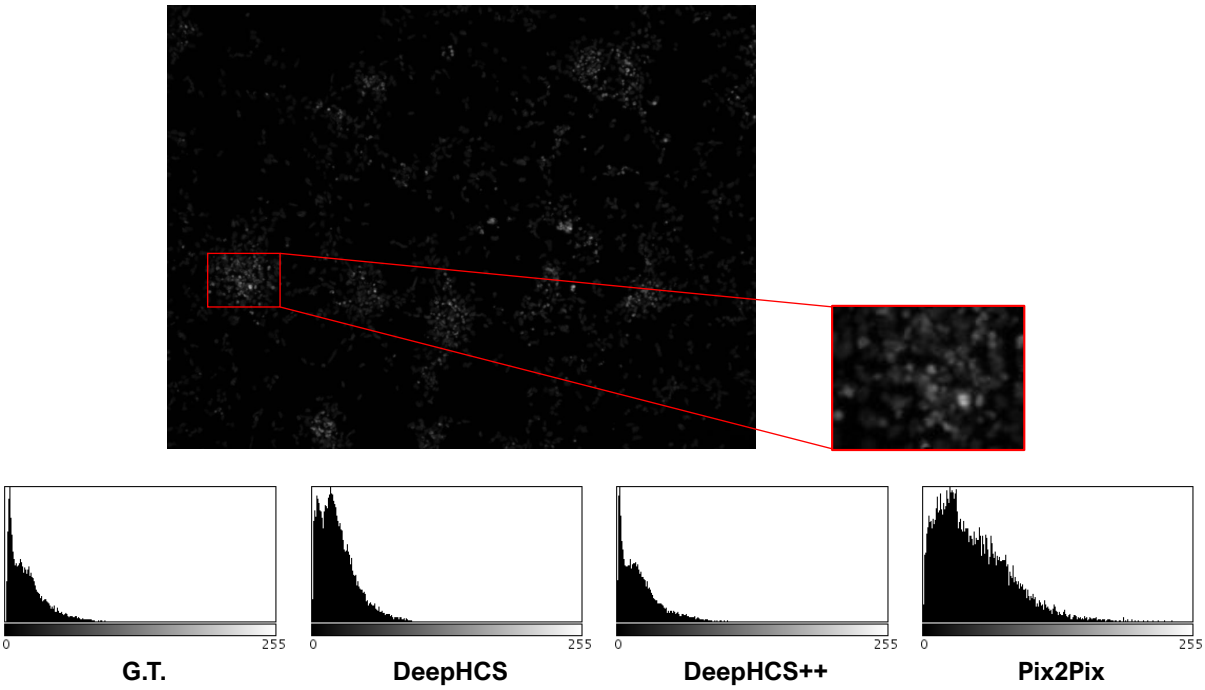


Figure 28: Comparison of the pixel intensity histogram of the resulting images (measured in the red box of the 35th image in Case 1 of Dataset1). The adversarial loss in DeepHCS^{++} improves the intensity distribution of the resulting image closer to that of the ground-truth image.

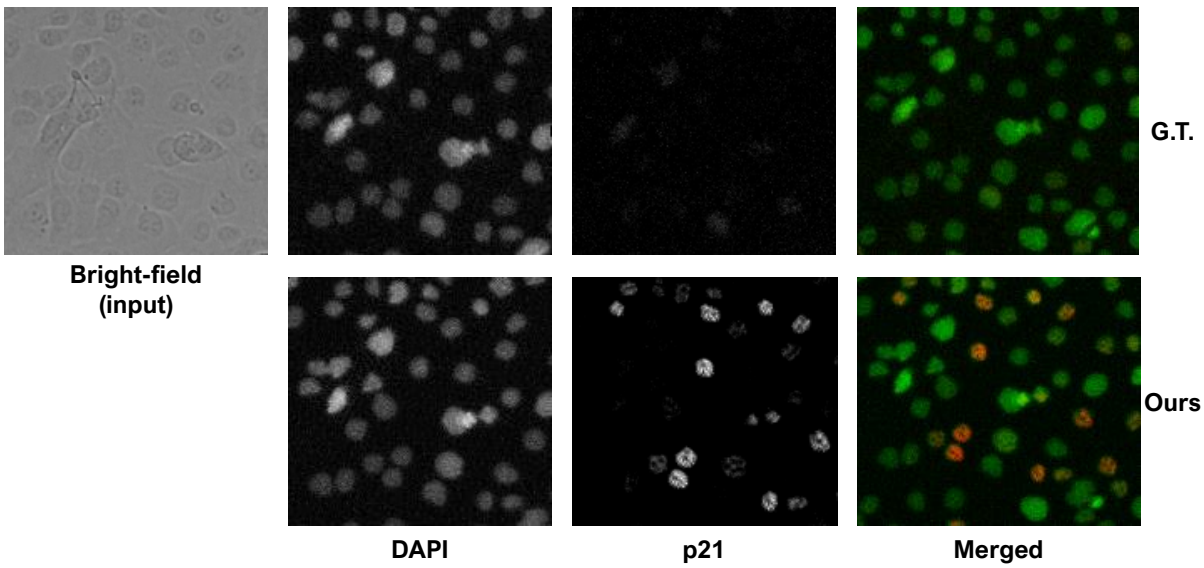
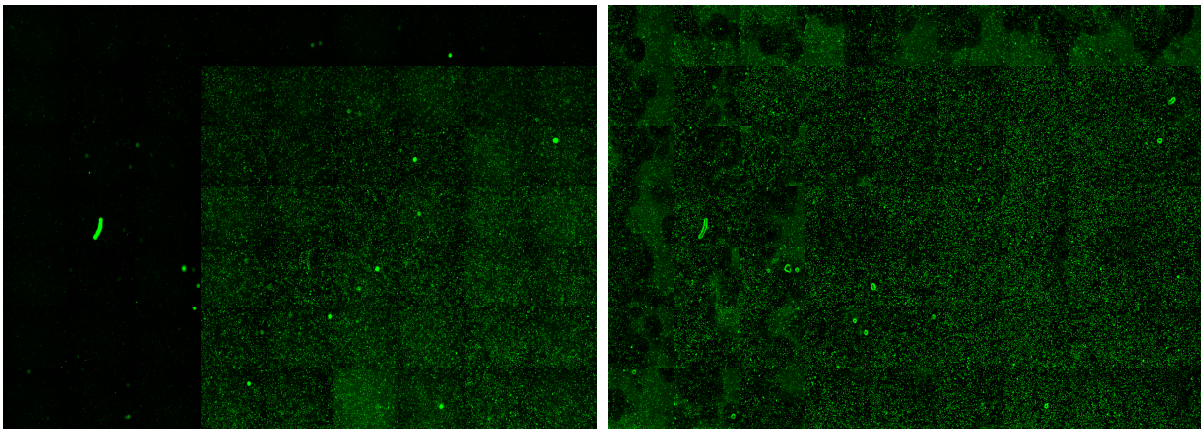


Figure 29: Comparison of ground-truth non-TGF-β images and DeepHCS^{++}-generated images. The visual quality of generated fluorescence images from DeepHCS^{++} is acceptable for DAPI images, whereas p21 fluorescence images are incorrectly translated (showing many false positives). Because p21 protein can only be measured by staining cells with specific antibody, DeepHCS^{++} should not generate stained cells as shown in the bottom row. DAPI (green) and p21 (red) images are overlaid to show whether p21 is visible or not (rightmost column).



(a) p21 fluorescence images in G.T. (b) p21 fluorescence images in ours

Figure 30: Overview of p21 measurement in a well plate. Field-01 image (the center among the nine images, see Fig. **??**b) of each well is visualized. (a) Ground-truth. First row and left three columns are the cases of non-TGF-β and TGF-β+Galunisertib, therefore, p21 is not measured; (b) Result of DeepHCS^{++}. Since our method is relying on cell's morphological structures, p21 is equally measured in every well, which is an incorrect translation.

of the canonical responses to TGF-β treatment in cells is transcriptional activation of the target genes such as p21 [62], which we believe can only be measured through proper protein staining. To validate this hypothesis, we ran another test of translating p21-stained fluorescence images directly from bright-field images using DeepHCS^{++}. We used the same Dataset3 but it was stained with anti-p21 staining on nuclear and imaged with the same device setting introduced in Section 5.2. As in the EMT dataset experiment, the GBM-data pre-trained model was used for transfer learning on the p21 dataset. As shown in Figures 29 and 30, p21 protein was measured only in the region of TGF-β treatment. Although DAPI fluorescence images were correctly reconstructed using DeepHCS^{++}, the level of p21 protein of the generated p21 fluorescence images does not correlate with that of the real p21 images. This result confirms that the potential application of DeepHCS^{++} may only be limited to translation of cell structures from bright-field to fluorescence images.

Another limitation of DeepHCS^{++} is that the algorithm is memory-bound because the size of network depends on the number of biomarkers (i.e., an additional decoder in the TN and a new RN are required to generate the fluorescence image for a new biomarker). In addition, our current network depends on the magnification of the imaging process, and therefore, a new network should be trained from scratch to change the magnification of the input and output images.

Despite these limitations, we believe that DeepHCS^{++} provides new insights into label-free cell image analysis and will facilitate future analytical advances of high content analysis.

5.6 Summary

In this paper, we introduced DeepHCS^{++}, a novel deep end-to-end convolution neural network for generating three different fluorescence images directly from bright-field images. By leveraging multi-task learning, the DeepHCS^{++} can simultaneously generate three different fluorescence images, which are crucial for the analysis of the drug response of cancer patients.

Furthermore, adversarial learning helps DeepHCS^{++} produce realistic fluorescence images. We demonstrated that DeepHCS^{++} has a potential to reduce the laborious preparation process and improve the throughput of large-scale image-based drug screening process using deep learning. We also showed that DeepHCS^{++} can convert images from various cell types and formats, including time-lapsed live bright-field cell images, to fluorescence images without any biomarkers. In the future, we plan to apply DeepHCS^{++} to more different biomarkers and assess the efficacy of the method. Building a prototype in-silico HCS system using our method is another possible long-term future work.

VI Structure-Aware Microscopy Image Translation using Autoencoder for Cell Profile Feature Maps

6.1 Introduction

HCS has demonstrated their effectiveness in precision medicine consisting of readouts of various drug response to patient-derived cell cultures, and using high-throughput imaging and automatic image analysis to measure changes in the whole cells, such as the width, length, and shape of cells. In general, in order to facilitate the analysis of individual cells, these cells are stained using biomarkers that can sufficiently express the nucleus and cytoplasm of cells, and imaged under the transmitted-light microscope to obtain multiple fluorescence images corresponding to various biomarkers. The acquisition of different modalities of fluorescence images through a conventional method is time-consuming and laborious progress. In order to solve this problem, the method capable of generating various fluorescence images directly from a bright-field image through the previous method, DeepHCS^{++}, was proposed. DeepHCS^{++} effectively avoided the time-consuming and laborious tissue preparation process for generating biomarkers while providing accurate image analysis results. DeepHCS^{++} enables drug response and individual cell analysis by creating multiple modalities of fluorescence images similar to that of corresponding real fluorescence images. Even so, the size of the network increases proportionally when the type of fluorescence image required increases. This is the fatal limitation of GPUs which use limited memory compared to CPUs. Furthermore, if the entire cell shape needs to be translated, such as cytoplasm fluorescence images, a faint representation of the entire cell shape in the bright-field image will limit the complete conversion effort (Fig. 31).

Based on this observation, we propose an efficient structure-aware microscopy image translation (SAMIT) for reducing the number of trainable parameters and generating more realistic fluorescence images rather than previous works. In SAMIT, the convolutional autoencoder (CAE) [63] is proposed for multi-task learning with the translation network. This autoencoder can be used to learn a compressed representation of input data being composed of one encoder and decoder architecture. The encoder compresses the input, and the decoder attempts to reconstruct the input from the compressed information provided by the encoder which can be used as the data preparation technique to perform feature extraction on raw data. Based on these advantages, various fields are still employed, such as information retrieval [64], anomaly detection [65–68], image processing [69–72], and drug discovery [73].

Not only powerful performance regarding feature extraction, but also simple network architecture, the autoencoder is applied to the SAMIT to generate feature maps with rich textural information.

The contributions of the SAMIT can be summarized as follows:

- By employing the autoencoder as multi-task learning, the textural information in feature maps are sufficient, and the performance was better than in previous work.

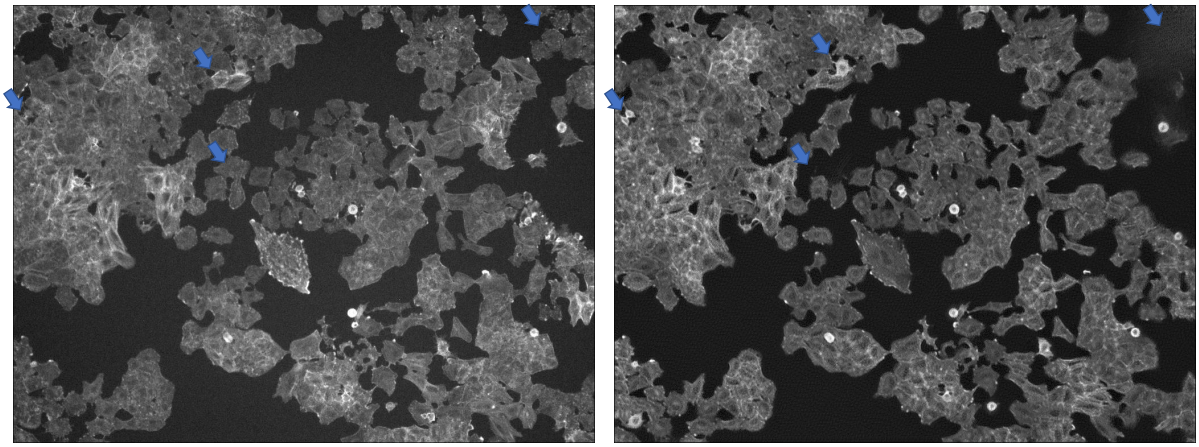


Figure 31: The previous work, DeepHCS^{++}, achieved the high performance on translated fluorescence images and the individual cell-related analysis. However, some cells can not reconstructed (blue arrows) since the texture on some cells highly blurred and fainted.

- Employing the feature mixture module with a cyclic consistency helps the SAMIT to reduce the size of the network (that is, training parameters).

We evaluated the accuracy of the SAMIT using widely-applied image quality metrics, PSNR and SSIM, and qualitatively validated our results versus the real cytoplasm fluorescence images using error map visualization and R^2 correlation.

6.2 Proposed Method: SAMIT

As with DeepHCS^{++}, the structure-aware microscopy image translation (SAMIT) employed multi-task learning to translate cytoplasm fluorescence images from bright-field images. The SAMIT is built upon two separate parts; the translation part, which translates bright-field images into corresponding fluorescence images and the feature mixture part, which improves the translated fluorescence images by leveraging feature maps from two preceding decoders. In the section below, each part is explained in detail.

Architecture Overview

The SAMIT consists of two different parts (Fig. 32). The first part of the SAMIT includes one encoder and two decoders; this encoder creates shared feature maps that have high-level compressed information, and then each decoder takes these shared feature maps to generate corresponding translated fluorescence images. The first decoder (Translator) handles the shared feature maps to generate cytoplasm fluorescence images, while the second decoder (Autoencoder) deals with the same shared feature maps to generate bright-field images. The one encoder and two decoders are reused by leveraging shared parameters on synthetic input images, $BF^{'}$, in the right side of Fig. 32.

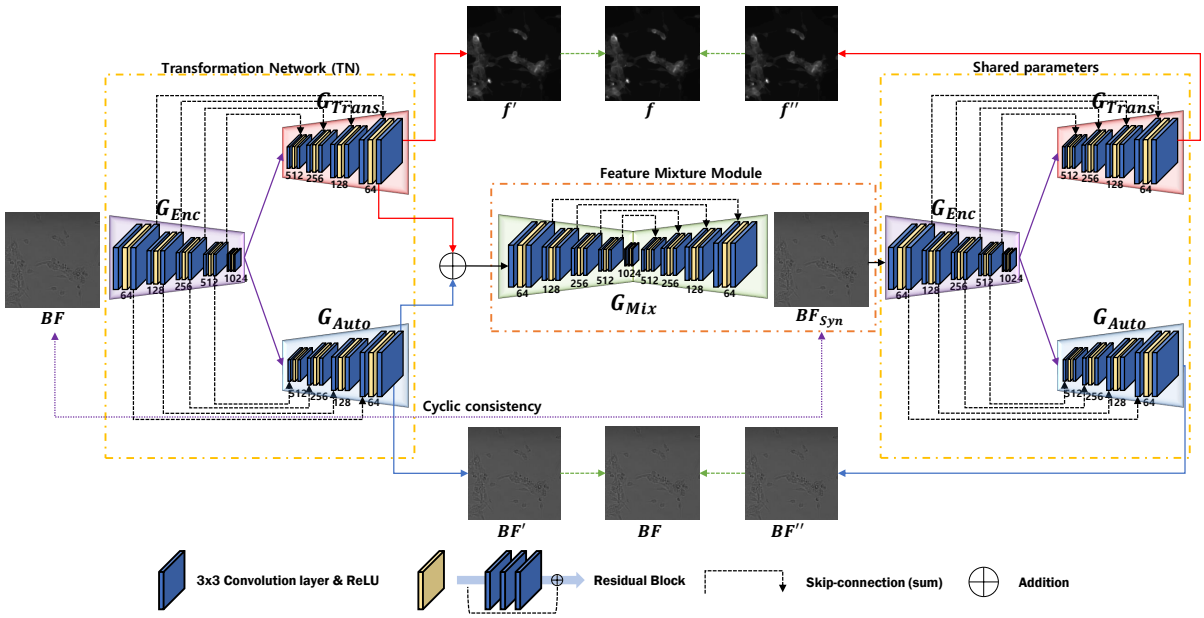


Figure 32: The SAMIT consists of two processes, first part is a translation part (one encoder and two decoders) in which the encoder generates a shared feature maps from input images and two decoders are operated as the translator and the autoencoder, respectively; second part is the feature mixture module (FMM) which takes the last feature maps from the translator and the autoencoder and generates synthetic input images which are connected with input images directly by a cyclic consistency. Convolution layers (blue) include ReLU as a non-linear function. Residual blocks (yellow) consist of three identical convolution layers. The filter size of all convolution layers used in the SAMIT is 3×3. Max-pooling layers and upscale (nearest-neighbor interpolation) layers are adapted to scale down and up after a set of "convolution-residual block-convolution" layers.

The second part of the SAMIT, the feature mixture module (FMM), is made up of U-Net [53] which generate synthetic input image, $BF^{'}$, allocated by the last feature maps of two preceding decoders. The generated synthetic input image, $BF^{'}$, is connected by a cycle consistency with input bright-field image, BF. Due to this cycle consistency between BF and $BF^{'}$, the quality of translated fluorescence image from the translator is significantly improved. In the following sections, each part of the SAMIT is explained in detail.

Translation Network with Autoencoder

The translation network with autoencoder is the first part of the SAMIT, consisting of one encoder and two identical decoders, which operate as the translator and the cell profile generator, respectively. The first network in the translation network is an encoder that takes a bright-field image as the input and generates shared feature maps (1024 feature maps), which have high-level compressed information for the next two identical decoders. In the training process, the encoder is trying to include the useful information from the bright-field images into shared feature maps so that two preceding decoders can accompany their corresponding task based on the high-level compressed information in these shared feature maps. As mentioned in Section V and the literature [49, 50], multi-task learning can obtain more necessary information in feature maps for each corresponding task. Hence, the translation network is built upon a multi-task network rather than an identical single-task network.

The second network in the translation network is comprised of two identical decoders that generate cytoplasm fluorescence images and bright-field images. These two identical decoders fully execute their corresponding tasks; the first decoder, G_E, for translating cytoplasm fluorescence images and the second decoder, G_A, for reconstructing bright-field images. The combination of the encoder, G_E, and the decoder, G_T, is considered the translation network (TN). In contrast, the combination of the encoder, G_{Enc}, and the decoder, G_A, is considered the autoencoder network (AN). In the flow of the AN, the textural information like the morphological feature of cells is almost recovered (Fig. 33b), and then this cell profile information is referenced to revise and improve synthetic fluorescence images generated by the TN with subsequent step network. In DeepHCS^{++}, generated fluorescence images with respect to apoptosis, DAPI, and cytoplasm are verified to be used for the drug response and analysis of individual cells. Nevertheless, the limitation still exists. In the case of apoptosis and DAPI fluorescence images, the purpose of these fluorescence images is to determine the presence or absence of nuclei of cells and convert it accurately. For this purpose, it is added to the cytoplasmic fluorescence image along with textural information through the region of nuclei of cells that is relatively well expressed in the bright-field image. Each fluorescence image is successfully converted using textural information. In the case of cytoplasm fluorescence images, the purpose of these fluorescence images is to clearly translate the cytoplasm of cells and the presence of nuclei of cells to enable the analysis of morphological features of individual cells. However, unlike the case of the apoptosis and DAPI fluorescence image translation, when the cytoplasmic fluorescence image is converted, there is a

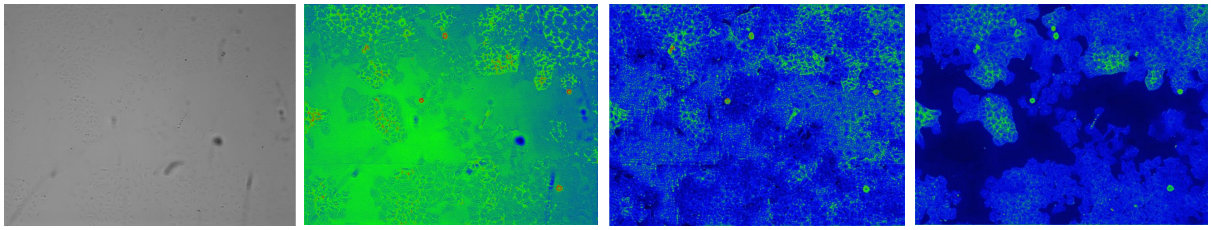


Figure 33: Examples of the last feature maps from the decoder on different tasks. (a) bright-field image as input image (b) is is feature map from the AN, and (c) feature map from the TN. (d) is feature map from the TN after applying FMM. To visualize more clearly, (b), (c) and (d) is summed feature map up along channel axis.

limitation in that the cytoplasm of the cell must be converted by guessing the shape of the cytoplasm using only the bright-field image's textural information even through multi-task learning has been employed. Depending on the characteristics of the fluorescence microscope image taken with light shining from below, the edge of the well is limited in some cases than the cells present in the center, which can degrade the quality of the converted cytoplasmic fluorescence image (Fig. 31).

In order to solve this problem, the AN is employed in the SAMIT and provides the textural information to the TN so that this cell profile information of the last feature maps from the AN decoder will be fused with the last feature maps from the decoder of the TN and then processed by the preceding network.

Feature Mixture Module

The feature mixture module (FMM) is the second part of the SAMIT (Fig. 32) and built upon U-Net [53], however, the other network consisting of the encoder and decoder architecture can be replaced with U-Net, which is employed as FMM to use less trainable parameters instead of deeper encoder-decoder architecture.

The FMM tasks the last feature maps from two decoders of the TN and AN and generates synthetic bright-field images, BF_{Syn}, fusing them to revise the shape of cells by connecting a cyclic consistency between the input image, BF, and the synthetic bright-field image, BF_{Syn}. The FMM is trained to accurately convert the fused feature maps into the bright-field images via a cyclic consistency. In FMM, the cyclic consistency is calculated to convert the fused feature maps into the synthetic bright-field image that is very same as the bright-field image and back-propagated to affect the previous two decoders so that unnecessary textural information (Fig. 33c) is eliminated, and the necessary one (Fig. 33b) is pointed out to restore the feature maps of the previous decoders. It is gradually learned so that only necessary textural information such as the cytoplasm of cells exists as much as possible (Fig. 33d).

Loss Function

For the TN and AN, the mean-absolute error (MAE) and the SSIM are used as the loss function to deal with the cell shapes and pixel intensity simultaneously. The MAE is defined as follows:

$$L_{MAE}(x,y) = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{24}$$

where x is the translated and reconstructed images of the TN and AN, respectively, and y is the target images for their corresponding tasks.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{25}$$

where μ_x and σ_x represent the mean and variance of image x, respectively, σ_{xy} represents the covariance of images x and y, and c_1 and c_2 are two constant variables for division stability. Based on Eq. 25, we can measure the degree of structural change in the image and recognize the difference between the two images based on luminance and contrast.

Based on Eq. 24 and 25, we define the loss function for the TN and AN by combining the two error measurements, as follows (α is empirically set to 0.8):

$$L_{TN} = ((1-\alpha)\cdot L_{MAE}(f^{'}, f) + \alpha \cdot L_{SSIM}(f^{'}, f)) + ((1-\alpha)\cdot L_{MAE}(f^{''}, f) + \alpha \cdot L_{SSIM}(f^{''}, f)) \tag{26}$$

$$\begin{aligned} L_{AN} = &((1-\alpha)\cdot L_{MAE}(BF^{'}, BF) + \alpha \cdot L_{SSIM}(BF^{'}, BF)) \\ &+ ((1-\alpha)\cdot L_{MAE}(BF^{''}, BF) + \alpha \cdot L_{SSIM}(BF^{''}, BF)) \end{aligned} \tag{27}$$

For the cyclic consistency between BF and BF_{Syn}, MAE is employed and defined as follows:

$$L_{cycle}(BF, BF_{Syn}) = \frac{1}{n}\sum_{i=1}^{n}\left|BF^i - BF_{Syn}^i\right| \tag{28}$$

The conditional adversarial loss was employed for the TN and AN to improve the quality of its results. Conditional GANs is defined as follows:

$$L_{CGAN}(G,D) = \min_G \max_D \mathbb{E}_{x,y}[log(D(x,y))] + \mathbb{E}_{x,z}[log(1 - D(x, G(x,z)))] \tag{29}$$

where x is a real input image, y is a real label image, and z is a generated image of the TN and AN. G, the TN and AN, tries to minimize Eq. 29, whereas D tries to maximize it.

The final loss functions for SAMIT are defined as follows:

$$L_{total} = L_{TN} + L_{AN} + 0.05 \cdot L_{CGAN} + L_{cycle} \tag{30}$$

Network implementation and Training process

The SAMIT consists of convolutional layers with non-linear activation functions to build up the TN and AN. We used a 3×3 convolutional layer with a stride of 1 and an ReLU function for the non-linear activation function. The zero-padding operation, “SAME" in TensorFlow, was used to

keep the spatial dimension of feature maps, and the number of filters was doubled each level from 64 up to 1024. After a set of "convolution - residual block - convolution" layers, max-pooling or nearest-neighbor interpolation (up-sampling) was used when feature maps were down-sampled or up-sampled. At the end of the TN and RN, a 1×1 convolutional layer with a stride of 1 was used to predict the final translated fluorescence image. With the exception of the encoder's last level, the two decoders and one encoder have an identical network. Skip-connections are connected between encoder and decoders via the same network level for recovering upper hierarchical information.

The discriminator consisted of five convolution layers; where the number of filters is doubled in every layer from 64 to 512 except in the last (5th) convolution layer, consisting of only one 4×4 filter. The first three convolutional layers used a 4×4 filter size and a stride of 2 instead of using a pooling layer to down-sample feature maps. The last two convolution layers used a 4×4 filter size and a stride of 1 to maintain the feature map's spatial dimension. After five convolution layers, a sigmoid function was used to produce the output of the discriminator's loss function.

All parameters of the SAMIT are trained based on Eq. 30 together. During the training, the following two data augmentation methods were used: (1) randomly flipping the images horizontally or vertically, and (2) randomly rotating the images by $90°, 180°$, or $270°$. We used a mini-batch of size 8; the Adam optimizer was used with a learning rate of 10^{-4}.

In the inference stage, the encoder (G_{Enc}) and decoder (G_{Trans}) are only deployed by taking bright-field images.

6.3 Results

In this experiment, the A549 lung adenocarcinoma cell line was chosen to validate our method. As mentioned in Section V, the columns of 1×7 wells in one half of a 384-well plate is administered with TGF-β at various dosage levels starting from 0.005 ng/ml; then the amount was increased three times per well. In the other half of the 384-well plate, TGF-β is not administered in half the remaining wells, whereas the galunisertib, an inhibitor, is administered to the last half at $10\mu/mol$. We used the training set including half of the A549 lung adenocarcinoma cell line bright-field images and their corresponding cytoplasm fluorescence images, in which each image is composed of 256×256 pixels.

To prove the performance of the SAMIT, we compared it with other methods in which single-task learning, is defined by one encoder (G_E) and one decoder (G_T) that can be considered as essentially pix2pix [1], although the depth of the network is different. The second is multi-task learning defined by the one encoder (G_E) and two decoders (G_T, G_A) together. The third is to prove the utility of the FMM defined by the translation network (G_E, G_T, G_A) and FMM. The fourth is the final network, as shown in Fig. 32 and named SAMIT.

In Fig. 34, the average PSNR of the third method is 22.1896, while the average PSNR of the second method is 21.7414. Besides, SAMIT is better than the third method by increasing from

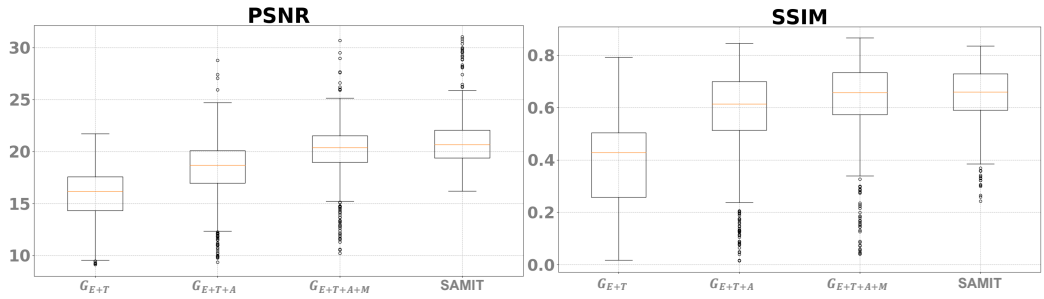


Figure 34: The ablation study of SAMIT with three different network using PSNR and SSIM.

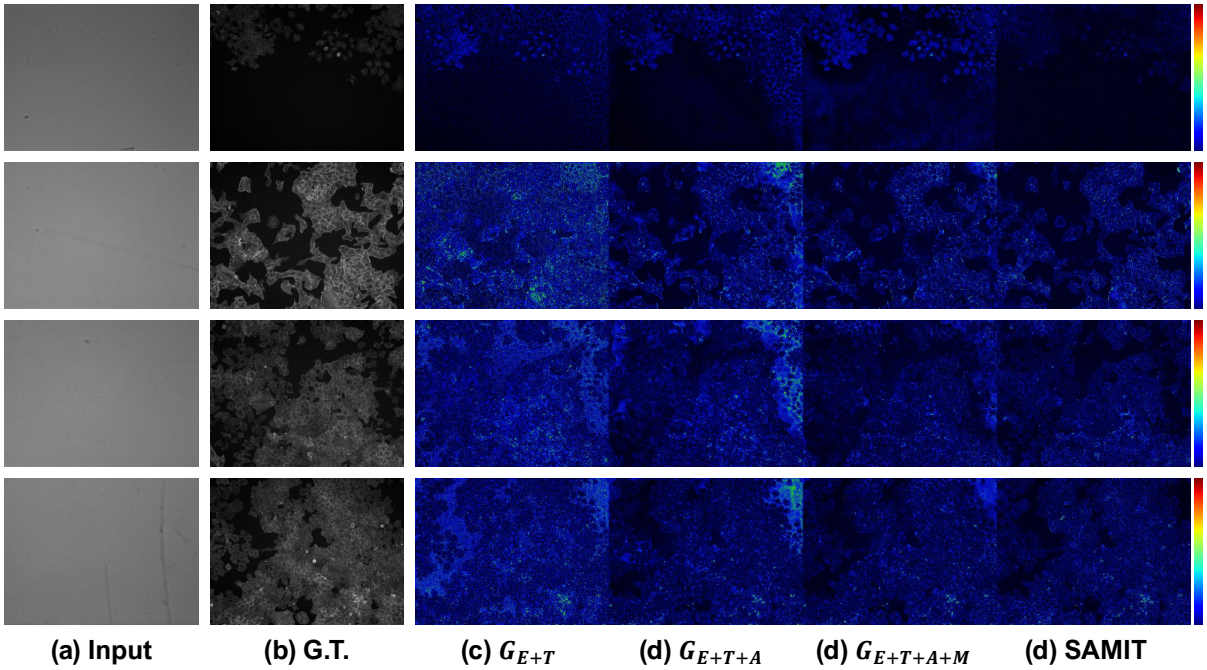


Figure 35: The qualitative comparison by visualizing translated fluorescence images by four different methods. The range of color bar is from 0.0 to 1.0.

22.1896 to 22.5738. It means that FMM and re-evaluation by shared parameters are proved to achieve performance gain. In the same way as the previous experiment, PSNR evaluation, we observed that FMM and re-evolution positively affect on incremental performance improvements in final translated images.

In Fig. 35, the fluorescence images converted by the aforementioned four methods were visualized, and qualitative comparison was conducted. Fig. 35a is the input images and Fig. 35b is the ground-truth of the cytoplasm fluorescence image. From Fig. 35c to Fig. 35f, it is the final translated fluorescence image of each method, and the difference was confirmed by applying a color map. The smaller the difference from the ground-truth of cytoplasm fluorescence image, the darker blue appears, and the larger the difference is, the redder it appears.

6.4 Discussion

In this section, we compare the SAMIT with DeepHCS^{++}, and discuss the strength and limitations of the SAMIT.

First of all, we will discuss the operation used for the input of FMM. Several methods can be proposed for fusing feature maps generated by different flows. The easiest and most commonly used method can be mainly divided into two categories: one is concatenation, and the other is addition. Concatenation has the advantage of being able to use feature maps created in different flows, but memory consumption are likely to occur during training as the number of feature map channels increases. On the other hand, since the Addition is processed after adding feature maps generated in different flows, it undergoes an unintended pre-processing phase. However, there is no increase in channels, so, problems such as memory consumption can be effectively handled when using a deep CNN. When comparing the translated fluorescence images by learning 50 epochs using the concatenation and addition of FMM's input feature maps, it is confirmed that there is no significant difference, as shown in Fig. 36. Not only the shape of the cell but also the luminance of the translated fluorescence images maintain almost the same quality, Verifying that it would be better to apply addition to construct an efficient network.

The second discussion is how meaningful the cytoplasm fluorescence images converted through the SAMIT are for individual cell analysis. It is proved that the SAMIT in Fig. 34 and 35 show improved performance over other methods, but the quantitative value may not be satisfactory when directly compared with the ground-truth of the cytoplasm fluorescence image. The main reason for this is that there are some artifacts/noises that fundamentally exist in the fluorescence microscope due to the photographing environment and settings. Thus, the translated fluorescence images appear relatively low in the quantitative numerical analysis based on pixel intensity.

Like Dataset3 in Section 5.4, we measured the average length, width, and width-to-length ratio of cells per well by performing segmentation of each cell using Cellpose [61] to validate a correlation between ground-truth and generated cytoplasm fluorescence images. In Fig. 37, The R^2 correlation of cell morphology means that the trend of cell morphological changes in ours are more similar to all cases with Growth Factor, Growth Factor+Inhibitor, and No Growth Factor rather than DeepHCS^{++} (Fig. 37). These graphs shows that the R^2 correlation of the cell morphological analysis of the cytoplasm fluorescence images through the SAMIT is highly improved compared to DeepHCS^{++} on the length, width, ratio width to length, as well as the variance of each morphological characteristic analysis value is significantly reduced rather than DeepHCS^{++}.

Furthermore, in order to analyze the morphological features of cells in detail, the change of morphological features of cells in the ground-truth and generated cytoplasm fluorescence images were analyzed for the growth factor and growth inhibitor by administering a small amount of the growth factor and no growth factor (Fig. 38). In the case of the cell length, the morphological

changes of cells administered with Growth Factors show almost the same trend as the real results in both the DeepHCS^{++} and SAMIT. In the other cases, there are slight differences between the DeepHCS^{++} and SAMIT. When an inhibitor is administered, SAMIT shows a graph trend more realistically similar to the real result than DeepHCS^{++}. In addition, it was confirmed that when the growth factor was not administered, there was little change in cells as shown in the result of the fluorescence image converted by SAMIT. In the case of the cell width, the SAMIT analysis results are very similar to the actual analysis results compared to DeepHCS^{++}. Compared to DeepHCS^{++}, which rapidly changed according to the dose of growth factor, SAMIT shows a smooth cell morphology change similar to the real analysis results. In the case of the cell width to length ratio, the length and width of cells in the fluorescence image generated by SAMIT result in a more similar analysis result than DeepHCS^{++}; thus, the cell width to length ratio of SAMIT show the results that are very similar to the analysis trend of real results.

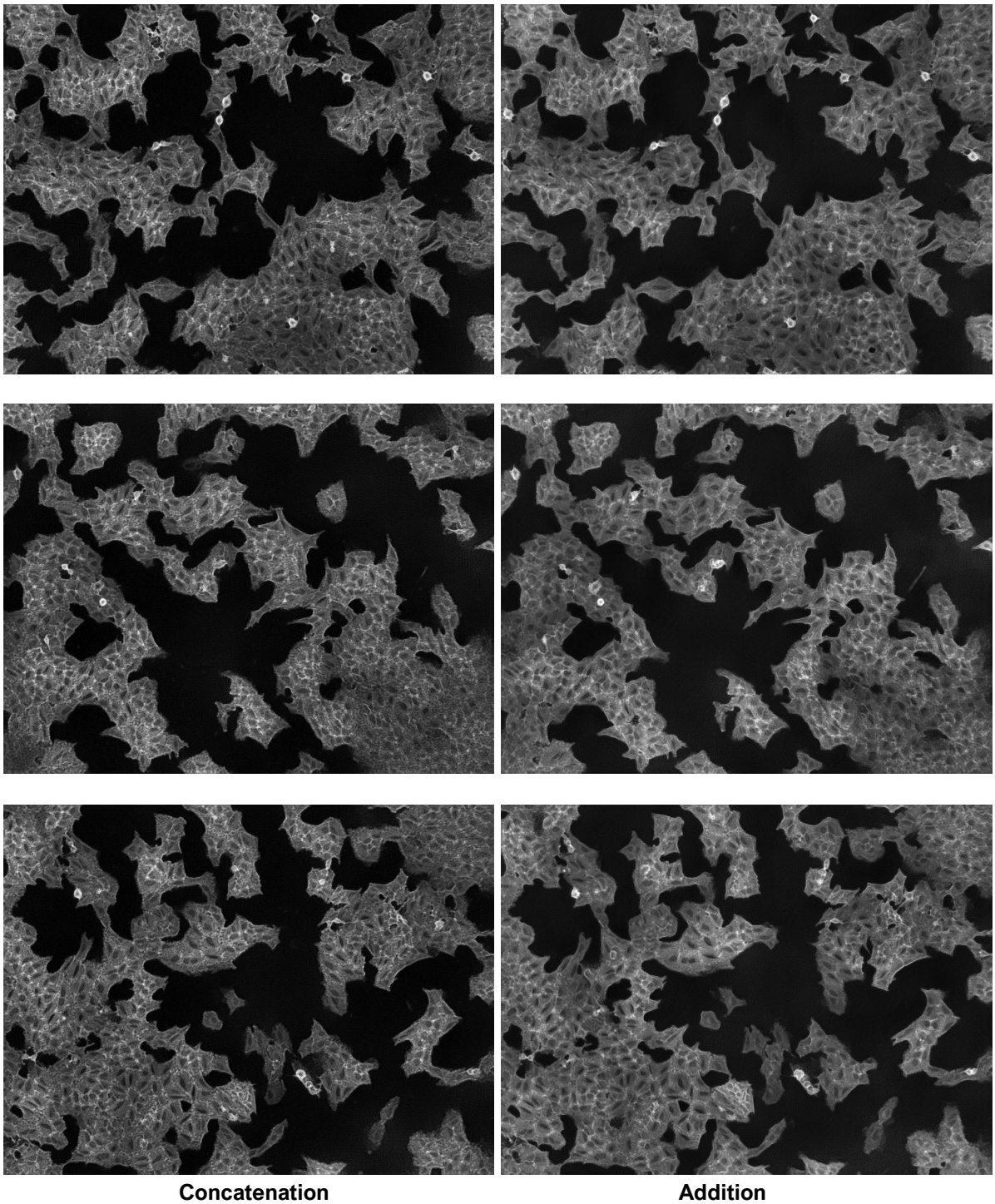


Figure 36: Examples of translated results affected by the concatenation and addition.

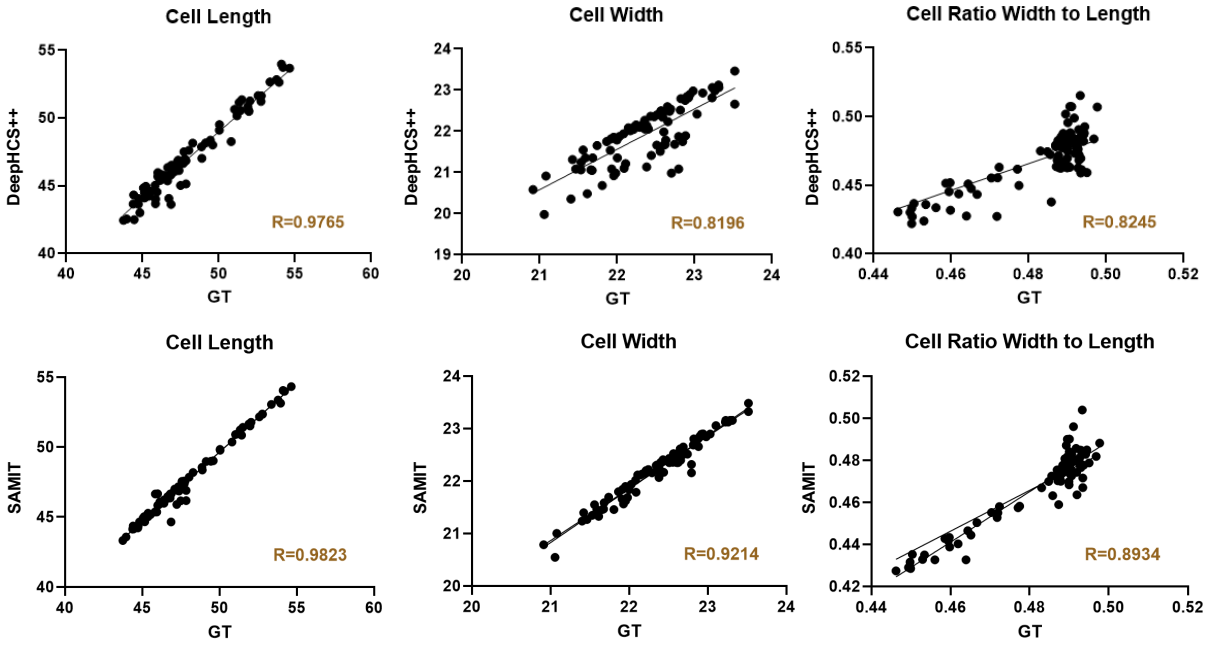


Figure 37: Comparison the SAMIT with DeepHCS^{++} in terms of the morphological features; the length, width, and ratio width to length of cells. The x-axis in all graphs represents measured values of the ground-truth. The y-axis represents DeepHCS^{++} and ours.

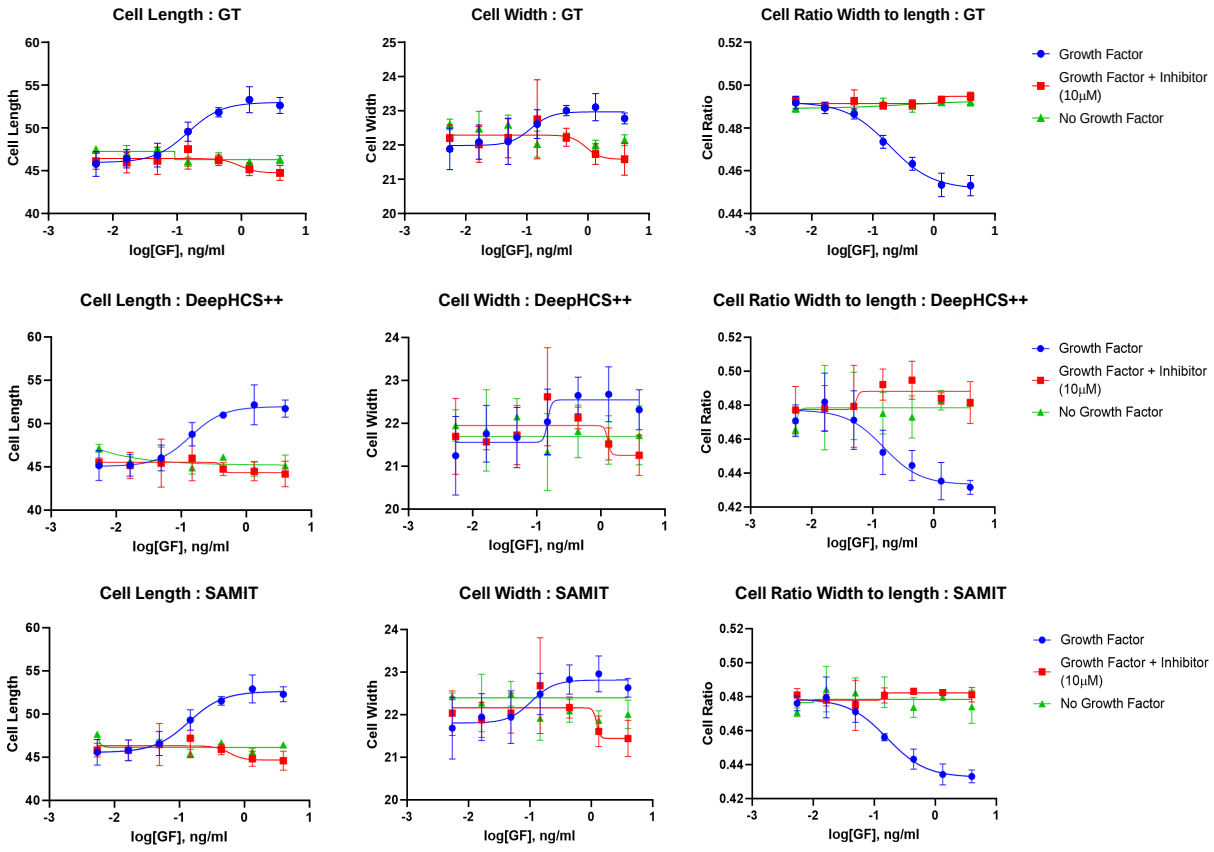


Figure 38: From top to bottom graphs, the fluorescence image of the ground-truth, DeepHCS^{++}, and the SAMIT. The y-axis represents the measured values with respect to the cell length, width, and ratio width to length and the x-axis represents the dosage (ng/ml) in log scale.

6.5 Summary

This section introduced SAMIT, an efficient deep end-to-end convolution neural network for generating corresponding fluorescence images directly from bright-field images. By leveraging multi-task learning, the SAMIT can simultaneously generate cytoplasm fluorescence images and synthetic bright-field images, which are used as reference information to recover the shape of cells.

Furthermore, the feature mixture module helps SAMIT effectively revise the shape of cells by using reference information from the fusion feature maps and generating improved synthetic fluorescence images. We demonstrated that SAMIT has the potential to reduce the number of trainable parameters than previous methods and improve the throughput of large-scale image-based drug screening process using deep learning. In the future, we plan to apply SAMIT to various biomarkers and assess the method's efficacy.

VII Conclusion

In this dissertation, I described a novel software-based high-content screening using several deep learning methods. The primary motivation of this research was to overcome the main drawback of inevitable steps in conventional the HCS pipeline, such as cell staining and cell fixation. The state-of-the-art deep learning technology was applied to effectively remove inevitable steps from the conventional HCS pipeline via proposed methods from bright-field microscopy images to corresponding fluorescence microscopy images, and concomitantly enable real-time drug reactivity analysis of cells. The results of drug reactivity and morphological analyses of individual cells in fluorescence microscopy images converted from single-task image conversion to multi-task image conversion methods are up to 97% when compared with actual fluorescence microscopy image analysis results. The results of the morphological analysis of individual cells was about 89%, showing a similar correlation. Additionally, I proposed a network that can effectively reduce the trainable parameter that is essential to add based on the increase of the number of fluorescence microscope images needed, the result was to create a better result than the proposed method using a lot of existing trainable parameters.

7.1 Summary of Dissertation Research

In order to find an appropriate anticancer treatment for individual patients, it is necessary to directly check drug response by administering anticancer treatments from several candidate groups directly to the patient's sample; thus, HCS is performed to analyze this effectively. In this process, cell staining and cell fixation are time-consuming and expensive operation. In Section IV, I present a novel end-to-end convolutional neural network, DeepHCS, to translate a bright-field image into DAPI fluorescence image in order to eliminate the cell fixation and cell staining. DeepHCS takes only one channel bright-field image to generate DAPI fluorescence image while related works [26, 52] required multi-focal bright-field images for fluorescence image translation. DeepHCS successfully translated the DAPI fluorescence image from a bright-field image and achieved a highly similar drug response for cell viability (up to 97% correlation). In addition, DeepHCS outperforms the state-of-the-art image translation method [1].

Even though DeepHCS was validated to replace the conventional HCS pipeline with software-based HCS analysis, it has a limitation when generating multiple corresponding fluorescence images simultaneously. In order to solve this problem, DeepHCS^{++} was proposed to translate a bright-field image into three corresponding fluorescence images: apoptosis, DAPI, and the cytoplasm of the cell in Section V. DeepHCS^{++} employed multi-task learning to generate multiple fluorescence images from a single bright-field image, and its performance was further improved than DeepHCS. In addition, time-lapse drug response analysis is possible, and its applicability to the GBM cells, as well as other cancer cells, is demonstrated, confirming that DeepHCS^{++} is not limited to specific cell types.

The more types of fluorescence images are required, the more trainable parameters are re-

quired since the refinement stage is present in the previously proposed method and increases proportionally to the number of trainable parameters, which can eventually cause memory consumption problems. To deal with this problem, in Section VI, SAMIT is proposed by utilizing the autoencoder and FMM to process generated feature maps required for the image conversion in order to resolve them effectively. Although additional trainable parameters are required in the training process by the SAMIT, for image translation, only the encoder and decoder are used in the inference process to dramatically reduce the size of the network. The transformed fluorescence images also showed improved results compared to the previously proposed methods.

7.2 Future work

From Section IV to Section VI, a novel end-to-end deep learning method is suggested for the fluorescence image translation. Each proposed method achieved a reasonable quality in terms of analyzing the drug response and individual cell morphological features. Nevertheless, there is still a task to be solved. Since all of the proposed methods essentially adopt supervised learning, the training dataset composition with the perfect pair of images is the top priority. In practice, constructing large-scale pairs of images in the related industry is time-consuming and expensive; therefore, constructing a training dataset for supervised learning is not an attractive process. Nowadays, to tackle the disadvantage of supervised learning, semi-supervised learning [74, 75] has been introduced to deal with labeled and unlabeled datasets for corresponding tasks. Alternatively, self-supervised learning [76–78], as a subset of unsupervised learning, is also introduced to learn representative features from large-scale unlabeled datasets without any human-annotated labels or image dataset pairs.

As future work, the state-of-the-art learning method mentioned above should be considered to break through the supervised learning limits. Expect research to be conducted on the image conversion technology that can flexibly respond to various datasets.

References

[1] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[2] G. Lee, J.-W. Oh, M.-S. Kang, N.-G. Her, M.-H. Kim, and W.-K. Jeong, "DeepHCS: Bright-Field to Fluorescence Microscopy Image Conversion Using Deep Learning for Label-Free High-Content Screening," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 335–343.

[3] S. Fox, S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, R. Khoury, and M. Biros, "High-throughput screening: update on practices and success," *Journal of biomolecular screening*, vol. 11, no. 7, pp. 864–869, 2006.

[4] H. Geerts, J. Q. Trojanowski, and V. Lee, "Drug discovery in neurodegenerative diseases." *Science of aging knowledge environment: SAGE KE*, vol. 2005, no. 6, pp. pe4–pe4, 2005.

[5] K. A. Giuliano, R. L. DeBiasio, R. T. Dunlay, A. Gough, J. M. Volosky, J. Zock, G. N. Pavlakis, and D. L. Taylor, "High-content screening: a new approach to easing key bottlenecks in the drug discovery process," *Journal of Biomolecular Screening*, vol. 2, no. 4, pp. 249–259, 1997.

[6] X. Long, L. Cleveland, and Y. L. Yao, "Automatic detection of unstained viable cells in bright field images using a support vector machine with an improved training procedure," *Computers in Biology and Medicine*, vol. 36, no. 4, pp. 339–362, 2006.

[7] F. Mualla, S. Schöll, B. Sommerfeldt, A. Maier, and J. Hornegger, "Automatic cell detection in bright-field microscope images using sift, random forests, and hierarchical clustering," *IEEE transactions on medical imaging*, vol. 32, no. 12, pp. 2274–2286, 2013.

[8] F. Buggenthin, C. Marr, M. Schwarzfischer, P. S. Hoppe, O. Hilsenbeck, T. Schroeder, and F. J. Theis, "An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy," *BMC bioinformatics*, vol. 14, no. 1, pp. 1–12, 2013.

[9] F. M. Iwamoto, A. R. Cooper, A. S. Reiner, L. Nayak, and L. E. Abrey, "Glioblastoma in the elderly," *Cancer*, vol. 115, no. 16, pp. 3758–3766, 2009.

[10] M. M. Jonczyk, J. Jean, R. Graham, and A. Chatterjee, "Trending towards safer breast cancer surgeries? examining acute complication rates from a 13-year nsqip analysis," *Cancers*, vol. 11, no. 2, p. 253, 2019.

[11] C. E. Quartararo, E. Reznik, A. C. DeCarvalho, T. Mikkelsen, and B. R. Stockwell, "High-Throughput Screening of Patient-Derived Cultures Reveals Potential for Precision Medicine in Glioblastoma," *ACS Medicinal Chemistry Letters*, vol. 6, no. 8, pp. 948–952, 2015.

[12] "Toward precision medicine in glioblastoma: The promise and the challenges," pp. 1051–1063, 2015.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[14] F. Gasparri, M. Mariani, F. Sola, and A. Galvani, "Quantification of the proliferation index of human dermal fibroblast cultures with the arrayscan™ high-content screening reader," *Journal of biomolecular screening*, vol. 9, no. 3, pp. 232–243, 2004.

[15] R. Ghosh, Y.-T. Chen, R. DeBiasio, R. DeBiasio, B. Conway, L. Minor, and K. Demarest, "Cell-based, high-content screen for receptor internalization, recycling and intracellular trafficking," *Biotechniques*, vol. 29, no. 1, pp. 170–175, 2000.

[16] G. J. Ding, P. A. Fischer, R. C. Boltz, J. A. Schmidt, J. J. Colaianne, A. Gough, R. A. Rubin, and D. K. Miller, "Characterization and quantitation of nf-κb nuclear translocation induced by interleukin-1 and tumor necrosis factor-α development and use of a high capacity fluorescence cytometric system," *Journal of Biological Chemistry*, vol. 273, no. 44, pp. 28 897–28 905, 1998.

[17] S. A. Haney, *High content screening: science, techniques and applications.* John Wiley & Sons, 2008.

[18] B. Neumann, T. Walter, J.-K. Hériché, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel *et al.*, "Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes," *Nature*, vol. 464, no. 7289, pp. 721–727, 2010.

[19] J. C. Simpson, B. Joggerst, V. Laketa, F. Verissimo, C. Cetin, H. Erfle, M. G. Bexiga, V. R. Singan, J.-K. Hériché, B. Neumann *et al.*, "Genome-wide rnai screening identifies human proteins with a regulatory function in the early secretory pathway," *Nature cell biology*, vol. 14, no. 7, pp. 764–774, 2012.

[20] M. V. Boland, M. K. Markey, and R. F. Murphy, "Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images," *Cytometry: The Journal of the International Society for Analytical Cytology*, vol. 33, no. 3, pp. 366–375, 1998.

[21] S. A. Haney, P. LaPan, J. Pan, and J. Zhang, "High-content screening moves to the front of the line," *Drug discovery today*, vol. 11, no. 19-20, pp. 889–894, 2006.

[22] R. Pepperkok and J. Ellenberg, "High-throughput fluorescence microscopy for systems biology," *Nature reviews Molecular cell biology*, vol. 7, no. 9, pp. 690–696, 2006.

[23] W. Buchser, M. Collins, T. Garyantes, R. Guha, S. Haney, V. Lemmon, Z. Li, and O. J. Trask, "Assay development guidelines for image-based high content screening, high content analysis and high content imaging," in *Assay guidance manual [Internet]*. Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2014.

[24] B. K. Shoichet, "Virtual screening of chemical libraries," *Nature*, vol. 432, no. 7019, pp. 862–865, 2004.

[25] D. J. E. Waibel, U. Tiemann, V. Lupperger, H. Semb, and C. Marr, "In-silico staining from bright-field and fluorescent images using deep learning," in *International Conference on Artificial Neural Networks*. Springer, 2019, pp. 184–186.

[26] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nature methods*, vol. 15, no. 11, pp. 917–920, 2018.

[27] Z. Darzynkiewicz, X. Li, and J. Gong, "Assays of cell viability: discrimination of cells dying by apoptosis," *Methods in cell biology*, vol. 41, pp. 15–38, 1994.

[28] J. Selinummi, P. Ruusuvuori, I. Podolsky, A. Ozinsky, E. Gold, O. Yli-Harja, A. Aderem, and I. Shmulevich, "Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images," *PloS one*, vol. 4, no. 10, p. e7497, 2009.

[29] R. Ali, M. Gooding, T. Szilágyi, B. Vojnovic, M. Christlieb, and M. Brady, "Automatic segmentation of adherent biological cell boundaries and nuclei from brightfield microscopy images," *Machine Vision and Applications*, vol. 23, no. 4, pp. 607–621, 2012.

[30] T. Tikkanen, P. Ruusuvuori, L. Latonen, and H. Huttunen, "Training based cell detection from bright-field microscope images," in *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2015, pp. 160–164.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[32] C. Cortes and V. Vapnik, "Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[33] K. Liimatainen, P. Ruusuvuori, L. Latonen, and H. Huttunen, "Supervised method for cell counting from bright field focus stacks," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on.* IEEE, 2016, pp. 391–394.

[34] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European Conference on Computer Vision.* Springer, 2016, pp. 702–716.

[35] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[36] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 1857–1865.

[37] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in neural information processing systems*, 2017, pp. 700–708.

[38] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 37–49.

[39] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[40] L. Han and Z. Yin, "Transferring microscopy image modalities with conditional generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 99–107.

[41] W. Jiang and Z. Yin, "Restoring the invisible details in differential interference contrast microscopy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2015, pp. 340–348.

[42] C. Fu, S. Lee, D. Joon Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Three dimensional fluorescence microscopy image synthesis and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 2221–2229.

[43] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning.* ACM, 2008, pp. 160–167.

[44] A. Sogaard and Y. Goldberg, "Deep multi-task learning with low level tasks supervised at lower layers," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 231–235.

[45] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 8599–8603.

[46] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7482–7491.

[47] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[48] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[49] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[50] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[51] R. Brent and L. Boucheron, "Deep learning to predict microscope images," *Nature methods*, vol. 15, no. 11, pp. 868–870, 2018.

[52] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee *et al.*, "In silico labeling: Predicting fluorescent labels in unlabeled images," *Cell*, vol. 173, no. 3, pp. 792–803, 2018.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[54] X. Li, G. Zhang, J. Wu, H. Xie, X. Lin, H. Qiao, H. Wang, and Q. Dai, "Unsupervised content-preserving image transformation for optical microscopy," *bioRxiv*, p. 848077, 2019.

[55] M. Boutros, F. Heigwer, and C. Laufer, "Microscopy-Based High-Content Screening," *Cell*, vol. 163, no. 6, pp. 1314–1325, 2015.

[56] T. M. Quan, D. G. Hilderbrand, and W.-K. Jeong, "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics," *arXiv preprint arXiv:1612.05360*, 2016.

[57] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[58] F. Piccinini, A. Kiss, and P. Horvath, "Celltracker (not only) for dummies," *Bioinformatics*, vol. 32, no. 6, pp. 955–957, 2015.

[59] Y. Hao, D. Baker, and P. ten Dijke, "TGF-β mediated epithelial-mesenchymal transition and cancer metastasis," *International journal of molecular sciences*, vol. 20, no. 11, p. 2767, 2019.

[60] E. Janda, K. Lehmann, I. Killisch, M. Jechlinger, M. Herzig, J. Downward, H. Beug, and S. Grunert, "Ras and TGF-β cooperatively regulate epithelial cell plasticity and metastasis: dissection of ras signaling pathways," *The Journal of cell biology*, vol. 156, no. 2, pp. 299–314, 2002.

[61] C. Stringer, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *bioRxiv*, 2020.

[62] N.-G. Her, J.-W. Oh, Y. J. Oh, S. Han, H. J. Cho, Y. Lee, G. H. Ryu, and D.-H. Nam, "Potent effect of the mdm2 inhibitor amg232 on suppression of glioblastoma stem cells," *Cell death & disease*, vol. 9, no. 8, pp. 1–12, 2018.

[63] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59.

[64] R. Salakhutdinov and G. Hinton, "Semantic hashing," *International Journal of Approximate Reasoning*, vol. 50, no. 7, pp. 969–978, 2009.

[65] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, 2014, pp. 4–11.

[66] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

[67] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 665–674.

[68] M. Ribeiro, A. E. Lazzaretti, and H. S. Lopes, "A study of deep convolutional auto-encoders for anomaly detection in videos," *Pattern Recognition Letters*, vol. 105, pp. 13–22, 2018.

[69] K. Cho, "Simple sparsification improves sparse denoising autoencoders in denoising highly corrupted images," in *International Conference on Machine Learning*, 2013, pp. 432–440.

[70] ——, "Boltzmann machines and denoising autoencoders for image denoising," *arXiv preprint arXiv:1301.3468*, 2013.

[71] K. Zeng, J. Yu, R. Wang, C. Li, and D. Tao, "Coupled deep autoencoder for single image super-resolution," *IEEE transactions on cybernetics*, vol. 47, no. 1, pp. 27–37, 2015.

[72] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2015.

[73] A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev *et al.*, "Deep learning enables rapid identification of potent ddr1 kinase inhibitors," *Nature biotechnology*, vol. 37, no. 9, pp. 1038–1040, 2019.

[74] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[75] N. N. Pise and P. Kulkarni, "A survey of semi-supervised learning methods," in *2008 International Conference on Computational Intelligence and Security*, vol. 2, 2008, pp. 30–34.

[76] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *arXiv preprint arXiv:2011.00362*, 2020.

[77] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[78] X. Liu, F. Zhang, Z. Hou, Z. Wang, L. Mian, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," *arXiv preprint arXiv:2006.08218*, vol. 1, no. 2, 2020.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Professor Won-Ki Jeong (Koera University)for the continuous support throughout my study and research. Without his guidance and constant feedback, my PhD course and this dissertation would not have been achievable. During my graduate life, thank you so much for giving me the opportunity to experience many research projects. This has enabled us to conduct various studies on various data, and I will never forget the grace of your advice to constantly think productively.

I would also like to thank my committee members, Prof. SeYoung Chun (UNIST), Prof. Jae-Young Sim (UNIST), Prof. Sungahn Ko (UNIST), Prof. Kwang In Kim (UNIST) for their encouragement, insightful comments toward this work. Based on productive comments, my dissertation has been very helpful in improving to the end.

In gratitude, I can not miss two guys, Jeong-Woo Oh and Nam-Gu Her in Samsung Medical Center, because my research works would have been limited without data generation and constructive feedback.

I am very grateful to our HVCL lab mates; Tran Minh Quan, Sumin Hong, Woohyuk Choi, Jungmin Moon, Junyoung Choi, Kanggen Lee, Haejin Jeong, Hyungjoon Jang, ChanMin Park, Hyeonsoo Lee, Nguyen Tuan Khoa, Jing Wei Tan, Tran Anh Tuan, Sungduk Cho and YeaIn Lee for their collaboration and support in various research areas related to this dissertation.

In addition, my brilliant friends and colleagues inspired me over the many years for making my experience in graduate school exciting and fun.

Last but not least, I would like to express my deepest gratitude to my family, younger sister and my parents. This dissertation would not have been possible without their warm love, continued patience, and endless support.

Finally, I want to give this pleasure to my girlfriend (Mi-Yeong Kim) who has always been with me so that I can successfully complete my Ph.D. program through constant encouragement and support from me during my not-so-short graduate life.